

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p3\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 18:47:19 ; Search time 1241 Seconds

(without alignments)  
8207.882 million cell updates/sec

Title: US-09-880-711-328

Perfect score: 350

Sequence: 1 atgagccacagctacccca.....ctgagagcttcacataaaaa 350

Scoring table: IDENTITY\_NUCDX

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4105280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl : \*

1: gb\_ba : \*

2: gb\_hlg : \*

3: gb\_in : \*

4: gb\_om : \*

5: gb\_ov : \*

6: gb\_pal : \*

7: gb\_ph : \*

8: gb\_pl : \*

9: gb\_pr : \*

10: gb\_ro : \*

11: gb\_sy : \*

12: gb\_un : \*

13: gb\_vl : \*

14: gb\_vl : \*

15: em\_ba : \*

16: em\_fun : \*

17: em\_hum : \*

18: em\_in : \*

19: em\_mu : \*

20: em\_om : \*

21: em\_of : \*

22: em\_ov : \*

23: em\_pat : \*

24: em\_ph : \*

25: em\_pl : \*

26: em\_ro : \*

27: em\_sts : \*

28: em\_un : \*

29: em\_vl : \*

30: em\_hlg\_hum : \*

31: em\_hlg\_inv : \*

32: em\_hlg\_other : \*

33: em\_hlg\_mus : \*

34: em\_hlg\_pln : \*

35: em\_hlg\_rtd : \*

36: em\_hlg\_mam : \*

37: em\_hlg\_vrt : \*

38: em\_sy : \*

39: em\_hlgo\_hum : \*

40: em\_hlgo\_mus : \*

41: em\_hlgo\_other : \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201.2	57.5	1315	5	BC003608
2	199.6	57.0	1326	9	D84109
3	171.6	49.0	2406	10	BC011288
4	169.4	48.4	1587	10	AF148511
5	167.8	47.5	1124	10	BC030397
6	150.4	43.0	209598	2	AC099473
7	142.8	40.8	1594	9	D84110
8	141.8	40.5	778	9	D84111
9	141.8	40.5	1388	9	D84107
10	141.4	40.4	61962	2	AC123616
11	141.4	40.4	262432	2	AC127551
12	124.2	35.5	1995	5	AK057533
13	124.2	35.5	126753	2	AF252826
14	124.2	35.5	127677	9	AC102945
15	124.2	35.5	155926	9	AC011204
16	117.8	33.7	1474	5	D84108
17	95.2	27.2	113320	2	AC122575
18	72	20.6	93263	2	AC103070
19	64.8	18.5	60155	5	AC083883
20	63.6	18.2	92143	9	HS608E8
21	63.4	18.1	37239	9	AC096549
22	62.6	17.9	174192	8	AF073629
23	62.6	17.5	235593	2	AC110534
24	62.6	17.5	235593	2	AL772296
25	62.6	17.9	293387	2	AL807802
26	62.2	17.8	265938	2	AC127247
27	61.6	17.6	125020	9	AF429315
28	61.4	17.5	158836	2	AC106126
29	61.4	17.5	170523	2	AC053526
30	61.4	17.5	170821	10	AL645626
31	61.4	17.5	177782	9	AC117532
32	61.4	17.5	178105	9	AC073057
33	61.4	17.5	189648	2	AC118994
34	61.2	17.5	152019	2	AC129992
35	61	17.4	179714	9	AC092555
36	61	17.4	184696	9	AC112507
37	60.8	17.4	7218	6	166494
38	60.8	17.4	135721	2	AC119606
39	60.5	17.3	39601	9	AC005581
40	60.6	17.3	131903	9	HS360E18
41	60.6	17.3	139843	2	AL158145
42	60.6	17.3	160065	2	AC023329
43	60.2	17.2	40468	3	AF040654
44	60.2	17.2	139258	9	HSDA60B16
45	60	17.1	2481	9	H0MNAK1

#### ALIGNMENTS

RESULT 1

BC003608 1315 bp mRNA linear p81 12-JUL-2001

LOCUS Homo sapiens, KNA-binding protein gene with multiple splicing, clone MGC:1257 IMAGE:3505124, mRNA, complete cds.

DEFINITION

ACCESSION BC003608

VERSION BC003608.1 GI:13177643

KEYWORDS

SOURCE MGC.

ORGANISM Homo sapiens.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS Strausberg, R.

TITLE Direct Submission







JOURNAL Submitted (04-MAY-1999) Zoology, University of Texas at Austin,  
C0900, Austin, TX 78712-1064, USA  
Location/Qualifiers

FEATURES  
SOURCE  
1.1587  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/tissue\_type="heart"  
/dev\_stage="adult"  
272..865  
/note="RRM-type RNA-binding protein"  
/codon\_start=1  
/product="hermes"  
/protein\_id="A039515.1"  
/db\_xref="GI:5081739"  
/translation="MNGGKAKEKENTPSPANTLOEEVETLTVSGILPDIKPREYLLF  
RPFKGEESLITKIKSKOPVGFVSFDSRSRAEAKANLNGIRDPDEIPOTLRFEKKN  
TKMAKNKLVGTNPSTPLPNTVPHSVAREPVELTYPALYPSSEVMAPYPLAPLAP  
ALPPPAATYTPASLHAQMKWIPPSSEATSGMKSRQFC"

CDS  
BASE COUNT 371 a 460 c 370 g 373 t 13 others

ORIGIN

Query Match  
Best Local Similarity 80.6%; Pred. No. 8.1e-18;  
Matches 170; Conservative 15; Mismatches 26; Indels 0; Gaps 0;

1 ATGAGCTCAGTACTGCTTACCCAGTAGAGTGGCCCGGTACCTC 60  
DB 669 ATGAGCTCAGTACTGCTTACCCAGTAGAGTGGCCCGGTACCTC 728

61 TGTACCCAGGAGTATGAGCCCTGCTTCTCTCCGCGCTTACCTACCCGCTT 120  
DB 729 TGTACCCAGGAGTATGAGCCCTGCTTCTCTCCGCGCTTACCTACCCGCTT 788

121 CACTGATGCCAGAGGAGCCTGTATACATGCTAGAGNNNNACNCTGACNNNTGAGANAC 180  
DB 789 CACTGATGCCAGAGGAGCCTGTATACATGCTAGAGNNNNACNCTGACNNNTGAGANAC 848

181 TMCCTACGACATGATGATNCCTGCTGCG 211  
DB 849 CACGGCAGTTCTGCTGAGGCTGTCTCGGG 879

RESULT 5  
LOCUS BC030397 1124 bp mRNA linear ROD 07-AUG-2002  
DEFINITION Mus musculus, RNA binding protein gene with multiple splicing,  
clone MGC:40656 IMAGE:4225078, mRNA, complete cds.  
ACCESSION BC030397  
VERSION BC030397.1 GI:20988888  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1124)  
AUTHORS Strausberg, R.  
JOURNAL Direct Submission  
Submitted (07-MAY-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-rt@mail.nih.gov](mailto:cgapbs-rt@mail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Altier, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B.,

Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,  
Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantip, S., Thomas, P.J., Touchman, J.W.,  
Tsungue, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRK Plate: 65 Row: b Column: 19  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Similarity but not  
identity to protein

FEATURES  
SOURCE  
1.1124  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="FVB/N"  
/clone="MGC:40656 IMAGE:4225078"  
/tissue\_type="Colon, normal, 5 month old male mouse."  
/clone\_id="NCI-CGAP\_CO24"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
284..877  
/codon\_start=1  
/product="RNA binding protein gene with multiple splicing"  
/protein\_id="AAH30397.1"  
/db\_xref="GI:20988889"  
/db\_xref="locusid:19663"  
/translation="MNGGKAKEKENTPSPANTLOEEVETLTVSGILPDIKPREYLLF  
RPFKGEESLITKIKSKOPVGFVSFDSRSRAEAKANLNGIRDPDEIPOTLRFEKKN  
TKMAKNKLVGTNPSTPLPNTVPHSVAREPVELTYPALYPSSEVMAPYPLAPLAP  
ALPPPAATYTPASLHAQMKWIPPSSEATSGMKSRQFC"

CDS  
BASE COUNT 232 a 385 c 283 g 224 t

ORIGIN

Query Match  
Best Local Similarity 47.9%; Score 167.8; DB 10; Length 1124;  
Matches 169; Conservative 15; Mismatches 27; Indels 0; Gaps 0;

1 ATGAGCTCAGTACTGCTTACCCAGTAGAGTGGCCCGGTACCTC 60  
DB 681 ATGAGCTCAGTACTGCTTACCCAGTAGAGTGGCCCGGTACCTC 740

61 TGTACCCAGGAGTATGAGCCCTGCTTCTCTCCGCGCTTACCTACCCGCTT 120  
DB 741 TGTACCCAGGAGTATGAGCCCTGCTTCTCTCCGCGCTTACCTACCCGCTT 800

121 CACTGATGCCAGAGGAGCCTGTATACATGCTAGAGNNNNACNCTGACNNNTGAGANAC 180  
DB 801 CACTGATGCCAGAGGAGCCTGTATACATGCTAGAGNNNNACNCTGACNNNTGAGANAC 860

181 TMCCTACGACATGATGATNCCTGCTGCG 211  
DB 861 CCGGCAGTTCTGCTGAGGCTGTCTCGGG 891

RESULT 6  
LOCUS AC099473 209598 bp DNA linear HTG 13-JUL-2002  
DEFINITION Rattus norvegicus clone CH230-149B17, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 26 unordered pieces.  
ACCESSION AC099473  
VERSION AC099473.4 GI:21730968  
KEYWORDS HTG; HTGS; PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 209598)  
REFERENCE Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,

Albrooks, S.L., Amaratingue, H.C., Ate, J.R., Ayele, M., Banks, T.,  
Barbata, J., Benton, J., Bimage, K., Blankenburg, K., Brannin, D.,  
Bouck, J., Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.K., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Day, L.A., Davis, C., Davy-Carroll, L., DeDertich, D.A.,  
DeLaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Eannath, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Fallis, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franzy, P.,  
Gaisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B.,  
Honsi, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,  
Karlssoon, B., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichteberg, O., Lieu, C., Liu, J., Liu, W., Louiseqad, H.,  
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mathney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,  
Mines, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, N., Neal, D., Nextson, J., Nextson, N., Nguyen, A., Nguyen, N.,  
Nguyen, M., Nickerson, E., Nwokenkwo, S., Cgah, M., Okunonu, G.,  
Ordunye, N., Oyiedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Reay, Y.,  
Rivas, M., Rojas, A., Rojthokan, I., Rolfe, M., Ruiz, S., Savary, C.,  
Scheiner, S., Scott, G., Shen, H., Shooshitari, N., Sisson, J.,  
Sodergren, E., Sonaike, T., Sparks, A., Stanley, II, Stone, II.,  
Sutton, A., Syatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
Williams, G., Williamson, A., Wleczek, R., Woodem, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

## TITLE

## REFERENCE

## AUTHORS

## JOURNAL

## JOURNAL

## REFERENCE

## AUTHORS

## JOURNAL

## COMMENT

On Jul 11, 2002 this sequence version replaced gi:17974456.

## JOURNAL

## REFERENCE

## AUTHORS

## JOURNAL

## JOURNAL

## REFERENCE

## AUTHORS

## JOURNAL

## JOURNAL

## REFERENCE

## AUTHORS

## JOURNAL

## JOURNAL

## REFERENCE

## AUTHORS

## JOURNAL

\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1415: contig of 1415 bp in length  
\* 1415: gap of unknown length  
\* 1516: contig of 1064 bp in length  
\* 2580: gap of unknown length  
\* 2679: contig of 1364 bp in length  
\* 2680: gap of unknown length  
\* 4043: contig of 1364 bp in length  
\* 4144: gap of unknown length  
\* 5512: contig of 1365 bp in length  
\* 5513: gap of unknown length  
\* 7063: contig of 1450 bp in length  
\* 7163: gap of unknown length  
\* 8188: gap of 1026 bp in length  
\* 8288: gap of unknown length  
\* 10423: contig of 2135 bp in length  
\* 10524: gap of unknown length  
\* 12951: contig of 2427 bp in length  
\* 13050: gap of unknown length  
\* 13051: contig of 4643 bp in length  
\* 17693: gap of unknown length  
\* 21225: contig of 3432 bp in length  
\* 21325: gap of unknown length  
\* 24504: contig of 3179 bp in length  
\* 24505: gap of unknown length  
\* 24604: gap of unknown length  
\* 27757: contig of 3153 bp in length  
\* 27758: gap of unknown length  
\* 32340: contig of 4483 bp in length  
\* 32341: gap of unknown length  
\* 32440: gap of unknown length  
\* 32441: contig of 4911 bp in length  
\* 37352: gap of unknown length  
\* 37451: gap of unknown length  
\* 41921: contig of 4470 bp in length  
\* 41922: gap of unknown length  
\* 50402: contig of 8381 bp in length  
\* 50502: gap of unknown length  
\* 50503: contig of 5874 bp in length  
\* 50537: gap of unknown length  
\* 56476: gap of unknown length  
\* 56477: contig of 9042 bp in length  
\* 65519: gap of unknown length  
\* 75037: contig of 9419 bp in length  
\* 75038: gap of unknown length  
\* 75138: gap of unknown length  
\* 75139: gap of unknown length  
\* 90877: contig of 15440 bp in length  
\* 90878: gap of unknown length  
\* 106037: contig of 15060 bp in length  
\* 106038: gap of unknown length  
\* 106137: gap of unknown length  
\* 125353: contig of 19216 bp in length  
\* 125354: gap of unknown length  
\* 125454: contig of 18937 bp in length  
\* 144391: gap of unknown length  
\* 144392: gap of unknown length  
\* 144901: contig of 18117 bp in length  
\* 14491: gap of unknown length  
\* 162808: gap of unknown length  
\* 162908: contig of 22219 bp in length  
\* 185126: gap of unknown length  
\* 185127: contig of 24372 bp in length.  
\* 185227: contig of 24372 bp in length.

## FEATURES

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

## SOURCE

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html)).  
\* NOTE: This is a working draft sequence. It currently

Query Match 43.0% Score 150.4: DB 2: Length 209598;  
Best Local Similarity 65.2% Pred. No. 1.2e-13;  
Matches 172: Conservative 21: Mismatches 71: Indels 0: Gaps 0;  
ORIGIN  
BASE COUNT 56344 a 46131 c 47953 g 53134 t 4036 others  
1..209598  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-149B17"

	exon	648..706 /gene="RBP-MS" /number=2 707..756 /gene="RBP-MS" /number=3 757..812 /gene="RBP-MS" /number=4 813..963 /gene="RBP-MS" /number=5 964..1094 /gene="RBP-MS" /number=7 1095..1164 /gene="RBP-MS" /number=12 1165..1370 /gene="RBP-MS" /number=13 1371..1471 /gene="RBP-MS" /number=14 1472..1594 /gene="RBP-MS" /number=15		
BASE COUNT	313 a	563 c	373 g	345 t
ORIGIN				
Query Match	40.8%, Score 142.8, DB 9, Length 1594;			
Best Local Similarity	64.9%, Pred. No. 4,1e-13;			
Matches 187,	Conservative 24;	Mismatches 72;	Indels 5;	Gaps 2;
Oy	1	ATGAGCTCAGACGTACTGCATTTCACCAGTAGCCGTGAAGTGTGGGCCCGTACCCCTC	60	
Db	964	ATGAGCTCAGACGTACTGCATTTCACCAGTAGCCGTGAAGTGTGGGCCCGTACCCCTC	1023	
Oy	61	TGTACCAGCGGAGTTAAGCGCTGTCTTCTCTCCTCGCGCTTTCACCTAACCCGGCTT	120	
Db	1024	TGTACCAGCGGAGTTAAGCGCTGTCTTCTCTCCTCGCGCTTTCACCTAACCCGGCTT	1080	
Oy	121	CACGTACATGCCAGAGGCGNCCTGTATACTGCTCAGGNNNACANCTCATCNNNTSGAANC	180	
Db	1081	CACGTACATGCCAGAGTGCGTGGCTCCCTCCCTCCGAGGCTACTTCTCAGGGGCTGGAAGT	1140	
Oy	181	TNCTNCAG--CACATGNATGATNGGCTNCCTGGGAATGNGCCTGNCACAGATCAACCTAC	238	
Db	1141	CCCCTCAGTTCTGCGATTAATACTTACCTTTCAGCATGTGCTACTAGAAGGACGAAACAATT	1200	
Oy	239	TGCAGACCANCNAAGGAGCTTNNCAGACAAGATGTATTANNAACCA	286	
b	1201	GCCCCCTTGTGGAGTACGCGCTAATAGAAGCCCTGATCGCAATAGAGA	1248	
RESULT 8				
D84111				
LOCUS	D84111	778 bp.	mRNA	linear PRI 14-APR-2000
DEFINITION	Homo sapiens mRNA for RBP-MS/type 5, partial cds.			
ACCESSION	D84111			
VERSION	D84111.1 GI:1669554			
KEYWORDS	RBP-MS; RBP-MS/type 5; alternative splicing; Werner syndrome.			
SOURCE	Homo sapiens cell_line:Hela CDNA to mRNA.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryotic; Metazoa; Chordata; Catarrhini; Hominiidae; Homo.			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (sites) Shimamoto,A., Kitao,S., Ichikawa,K., Suzuki,N., Yamabe,Y., Iinamura,O., Tokutake,Y., Satoh,M., Matsumoto,T., Kuromitsu,J., Kataoka,H., Sugawara,K., Sugawara,M., Sugimoto,M., Goto,M. and Furuchi,Y.			

TITLE A unique human gene that spans over 230 kb in the human chromosome 8p11-12 and codes multiple family proteins sharing RNA-binding motifs

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (20), 10913-10917 (1996)

MEPDLIN 97003106

REFERENCE 2 (bases 1 to 778)

AUTHORS Shimamoto, A.

TITLE Direct Submission

JOURNAL Submitted (21-MAR-1996) Akira Shimamoto, AGFNE Research Institute Co., Ltd.; 200 Kajiwara, Kamakura, Kanagawa 247, Japan (E-mail:akirashimo.po.ijnet.or.jp, Tel:0467-46-4571, Fax:0467-48-6595)

FEATURES

Source

1..778

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="8"

/map="8p11.2-12.1"

/cell\_line="Hela"

1..778

/gene="RBP-MS"

<1..298

/gene="RBP-MS"

/standard\_name="RNA-binding protein gene with multiple splicing"

/note="alternative splicing (see also D84107-D84111)"

/codon\_start=2

/product="RBP-MS/Type 5"

/protein\_id="BAI2225.1"

/db\_xref="GI:1669555"

/translation="PFKYEGLSLKLTSPKOPGVFSPDRSEAEAKNALNGIFDPEIPQTLRLFEFAKNTKMAKNKLVGTNPSTPLPNTVPOFLAREPLGANEHCRCQ"

<1..6

/gene="RBP-MS"

/number=2

7..56

/gene="RBP-MS"

/number=3

57..112

/gene="RBP-MS"

/number=4

113..263

/gene="RBP-MS"

/number=5

264..409

/gene="RBP-MS"

/number=6

410..540

/gene="RBP-MS"

/number=7

541..610

/gene="RBP-MS"

/number=12

611..711

/gene="RBP-MS"

/number=14

712..778

/gene="RBP-MS"

/number=15

BASE COUNT 180 a 211 c 181 g 206 t

ORIGIN

Query Match 40.5% Score 141.8; DB 9; Length 778;

Best Local Similarity 64.9%; Pred. No. 4,6e-13;

Matches 185; Conservative 24; Mismatches 72; Indels 4; Gaps 2;

OY 1 ATGAGCTCAGCTACTGACTTTACCCAGTAGCCCTGAAGTGTGGCCCGCTACCCCTC 60

|||||

DB 410 ATGAGCTCAGAGTCCCTGACTTTACCCAGTAGCCCTGAAGTGTGGCCCGCTACCCCTC 469

OY 61 TGTACCCAGGAGAGTATGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120

|||||

DB 470 TGTACCCAGGAGAGTATGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 526

OY 121 CACTGCATGCCAGAGCGCTGTATACCTCCTAGGCGNNNNCAACNCTGCACNNNTGGANMC 180

|||||

DB 527 CACTGCATGCCAGAGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 586

OY 181 TMTTCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240

|||||

DB 587 GCGGTACGTCG-TGCTGAATACATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 645

OY 241 CAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 285

|||||

DB 646 GTGGCTATTATGACATCTTCAGTGGTGCTACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 690

RESULT 9

D84107

LOCUS D84107 1368 bp mRNA linear PK1 14-APR-2000

DEFINITION Homo sapiens mRNA for RBP-MS/Type 1, complete cds.

ACCESSION D84107

VERSION D84107.1 GI:1669546

KEYWORDS RBP-MS; RBP-MS/Type 1; alternative splicing; Werner syndrome.

SOURCE Homo sapiens cell\_line:embryonic carcinoma, MCF14 CDNA to mRNA.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1388)

AUTHORS Shimamoto, A., Kitao, S., Ichikawa, K., Suzuki, N., Yamabe, Y., Imamura, O., Iokura, Y., Satoh, M., Matsunoto, T., Kurimitsu, J., Kato, K., Sugawara, K., Sugawara, H., Sugimoto, M., Goto, M. and Furukuchi, Y.

TITLE A unique human gene that spans over 230 kb in the human chromosome 8p11-12 and codes multiple family proteins sharing RNA-binding motifs

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (20), 10913-10917 (1996)

MEPDLIN 97008106

REFERENCE 2 (bases 1 to 1388)

AUTHORS Shimamoto, A.

TITLE Direct Submission

JOURNAL Submitted (21-MAR-1996) Akira Shimamoto, AGFNE Research Institute Co., Ltd.; 200 Kajiwara, Kamakura, Kanagawa 247, Japan (E-mail:akirashimo.po.ijnet.or.jp, Tel:0467-46-4571, Fax:0467-48-6595)

FEATURES

Source

1..1388

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="8"

/map="8p11.2-12.1"

/cell\_line="embryonic carcinoma, NEC14"

1..1388

/gene="RBP-MS"

1..647

/gene="RBP-MS"

/number=1

1..566

/gene="RBP-MS"

567..1157

/gene="RBP-MS"

/standard\_name="RNA-binding protein gene with multiple splicing"

/note="alternative splicing (see also D84107-D84111)"

/codon\_start=1

/product="RBP-MS/Type 1"

/protein\_id="BAI2225.1"

/db\_xref="GI:1669547"

/translation="MNNCGKAEKENTPSPRANLQEEVRLTFVSGHPLDIKPRELYLFPFKYEGLSLKLTSPKOPGVFSPDRSEAEAKNALNGIFDPEIPQTLRLFEFAKNTKMAKNKLVGTNPSTPLPNTVPOFLAREPLGANEHCRCQ"

648..706

/gene="RBP-MS"

/number=2

707..756

exon

COMMENT

12790	13486: contig of 697 bp in length
13487	13586: gap of 100 bp

```
* 13587 14284: contig of 698 bp in length
* 14285 14384: gap of 100 bp
* 14385 15071: contig of 687 bp in length
* 15072 15171: gap of 100 bp
* 15172 15872: contig of 701 bp in length
* 15873 15972: gap of 100 bp
* 15973 16662: contig of 690 bp in length
* 16663 16762: gap of 100 bp
* 16763 17452: contig of 690 bp in length
* 17453 17552: gap of 100 bp
* 17553 18261: contig of 709 bp in length
* 18262 18361: gap of 100 bp
* 18362 19037: contig of 676 bp in length
* 19038 19137: gap of 100 bp
* 19138 19840: contig of 703 bp in length
* 19841 19940: gap of 100 bp
* 19941 20649: contig of 709 bp in length
* 20650 20749: gap of 100 bp
* 20750 21449: contig of 700 bp in length
* 21450 21549: gap of 100 bp
* 21550 22220: contig of 671 bp in length
* 22221 22320: gap of 100 bp
* 22321 23010: contig of 690 bp in length
* 23011 23110: gap of 100 bp
* 23111 23802: contig of 692 bp in length
* 23803 23902: gap of 100 bp
* 23903 24592: contig of 690 bp in length
* 24593 24692: gap of 100 bp
* 24693 25403: contig of 711 bp in length
* 25404 25503: gap of 100 bp
* 25504 26201: contig of 698 bp in length
* 26202 26301: gap of 100 bp
* 26302 27005: contig of 704 bp in length
* 27006 27105: gap of 100 bp
* 27106 27772: contig of 667 bp in length
* 27773 27872: gap of 100 bp
* 27873 28565: contig of 693 bp in length
* 28566 28665: gap of 100 bp
* 28666 29362: contig of 697 bp in length
* 29363 29462: gap of 100 bp
* 29463 30154: contig of 692 bp in length
* 30155 30254: gap of 100 bp
* 30255 30972: contig of 718 bp in length
* 30973 31072: gap of 100 bp
* 31073 31756: contig of 684 bp in length
* 31757 31856: gap of 100 bp
* 31857 32525: contig of 669 bp in length
* 32526 32625: gap of 100 bp
* 32626 33318: contig of 693 bp in length
* 33319 33418: gap of 100 bp
* 33419 34129: contig of 711 bp in length
* 34130 34229: gap of 100 bp
* 34230 34935: contig of 706 bp in length
* 34936 35035: gap of 100 bp
* 35036 35732: contig of 697 bp in length
* 35733 35832: gap of 100 bp
* 35833 36538: contig of 706 bp in length
* 36539 36638: gap of 100 bp
* 36639 37325: contig of 687 bp in length
* 37326 37425: gap of 100 bp
* 37426 38119: contig of 694 bp in length
* 38120 38219: gap of 100 bp
* 38220 38900: contig of 681 bp in length
* 38901 39000: gap of 100 bp
* 39001 39704: contig of 704 bp in length
* 39705 39804: gap of 100 bp
* 39805 40499: contig of 695 bp in length
* 40500 40599: gap of 100 bp
* 40600 41284: contig of 685 bp in length
* 41285 41384: gap of 100 bp
* 41385 42075: contig of 691 bp in length
* 42076 42175: gap of 100 bp
* 42176 42868: contig of 693 bp in length
```

```
* 42869 42968: gap of 100 bp
* 42969 43673: contig of 705 bp in length
* 43674 43773: gap of 100 bp
* 43774 44483: contig of 710 bp in length
* 44484 44583: gap of 100 bp
* 44584 45279: contig of 696 bp in length
* 45280 45379: gap of 100 bp
* 45380 46065: contig of 686 bp in length
* 46066 46165: gap of 100 bp
* 46166 46858: contig of 693 bp in length
* 46859 46958: gap of 100 bp
* 46959 47645: contig of 687 bp in length
* 47646 47745: gap of 100 bp
* 47746 48418: contig of 673 bp in length
* 48419 48518: gap of 100 bp
* 48519 49234: contig of 716 bp in length
* 49235 49334: gap of 100 bp
* 49335 50021: contig of 687 bp in length
* 50022 50121: gap of 100 bp
* 50122 50814: contig of 693 bp in length
* 50815 50914: gap of 100 bp
* 50915 51630: contig of 716 bp in length
* 51631 51730: gap of 100 bp
* 51731 52430: contig of 700 bp in length
* 52431 52530: gap of 100 bp
* 52531 53243: contig of 713 bp in length
* 53244 53343: gap of 100 bp
* 53344 54043: contig of 700 bp in length
* 54044 54143: gap of 100 bp
* 54144 54841: contig of 698 bp in length
* 54842 54941: gap of 100 bp
```

Query Match 40.4% Score 141.4: DB 2: Length 61962:

Best Local Similarity 65.8% Pred. No. 2.9e-12:

Matches 173: Conservative 21: Mismatches 66: Indels 3: Gaps 1:

```
OY 1 ATGAGCTGCAAGTACTGTCACCTTATACCCAGTAGCCCTGAGTAGTGAGGCCCGTACCCTG 60
DB 53740 ATGAGCTGCAAGTACTGTCACCTTATACCCAGTAGCCCTGAGTAGTGAGGCCCGTACCCTG 53799

OY 61 TGTACCCAGCGAGGTTAGCGCGTCTGCTCTCTCTGCGGCTTTACCTACCGCGCTT 120
DB 53800 TGTACCCAGCGAGGTTAGCGCGTCTGCTCTCTCTGCGGCTTTACCTACCGCGCTT 53859

OY 121 CACTGCATGCGCAGGAGGCGCTGATATATAGCTCCTAGGNNNACNCGAGNNNTGAGNAC 180
DB 53860 CACTGCATGCGCAGGAGGCGCTGATATATAGCTCCTAGGCGCACCCTTCGACTACCCACCAACCGC 53919

OY 181 TNCCTNAGCAGCATGNNATGATNAGCTGCTGCTGATATGAGCTGAGGCTGACAGTACCTACTG 240
DB 53920 CCTTTGAGAGCGCGCT---GCCCGAGAGCAGCAGCATGAGTACTACCTACCGAGTACCA 53975

OY 241 CAGACCACNCAAGCAGCTTNG 263
DB 53977 GATTCACCTTCAGGAGTGAATG 53999

RESULT 11
LOCUS AC127551 262432 bp DNA linear HTG 11-AUG-2002
DEFINITION Mus musculus chromosome UNK clone RP24-547H23, WORKING DRAFT
ACCESSION AC127551
VERSION AC127551.5 G1:22138547
KEYWORDS HTG: HTGS_LIBRARY: HTGS_DRAFT,
SOURCE house mouse,
ORGANISM Mus musculus,
Fukariyola: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Pluriferia: Rodentia: Sciurognathii: Muridae: Murinae: Mus.
REFERENCE 1 (bases 1 to 262432)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BB0547H23
-----
Summary Statistics
-----
Sequencing vector: M13, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-primer ET, % of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 258204 bases at least Q40
Consensus quality: 259283 bases at least Q30
Consensus quality: 259803 bases at least Q20

```

NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

*	1	1128:	contig of 1128 bp in length
*	1129	1228:	gap of unknown length
*	1229	10138:	contig of 8910 bp in length
*	10139	10238:	gap of unknown length
*	10239	16401:	contig of 6163 bp in length
*	16402	16501:	gap of unknown length
*	16502	25546:	contig of 9045 bp in length
*	25547	25546:	gap of unknown length
*	25647	44334:	contig of 18588 bp in length
*	44335	44334:	gap of unknown length
*	44335	71554:	contig of 27120 bp in length
*	71555	71554:	gap of unknown length
*	162671	1626770:	contig of 9116 bp in length
*	162771	262019:	gap of unknown length
*	262020	262119:	contig of 99249 bp in length
*	262120	262132:	contig of 313 bp in length.

```

RES
source
    Location/Qualifiers
      1..262432
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /chromosome="UNK"
        /clone="RP24-547H23"
      1..1128
        /note="assembly_name:Contig5"
      1229..10138
        /note="assembly_name:Contig19"
      10239..16401
        /note="assembly_name:Contig20"
      16502..25546
        /note="assembly_name:Contig21"
      25647..44234
        /note="assembly_name:Contig22"
      4435..71454
        /note="assembly_name:Contig23"

```

Qy	1	ATAGGCTACAGACTCTCAGCTTATACCCAGTAGCGCTGAAGTGGGGCCCGTACCTCG	60
Db	97134	ATGAGCTACAGTACTCTGACTCTTATACCCAGTAGCGCTGAAGTGGGGCCCGTACCTCG	97193
Qy	61	TGTACCCAGCGAGTTAGACGCCGTCTCTTCTCTCTCTGCGGCTTTACACTACCCCGCTT	120
Db	97194	TGTACCCAGCGAGTTAGACGCCGTCTCTCTCTCTCTCTGCGGCTTTACACTACCCCGCTT	97253
Qy	121	CACGTCAATGCCAGGAGGCGNCTGTATACATGCTAGTNNNNCCNCTGACNNTTGAATC	180
Db	97254	CACGTCAATGCCAGGAGTAAATTATAGCCCATGCGACACTTCCACTACACCCACGACACAC	97313
Qy	181	TNCTTCAGACATGTGATGATNGCGTNCNTGGGAATNGCGCTGNCACAGATCAACTACTG	240
Db	97314	CCTTTCAGACCCGGT---CCCCAGAGCACACCAATGTGACTAACACCTTACCCAGCTAACA	97370
Qy	241	CAGACCAACNMAAGGAGCTTTNG	263
Db	97371	GATTCACCTCCAGAAAGTGAATG	97393

RESULT	12
LOCUS	AK057533
DEFINITION	Homo sapiens CDNA FLJ32971 f1s, clone TEST12008847.
ACCESSION	AK057533
VERSION	AK057533.1 GI:16553275
KEYWORDS	oligo capping; f1s (full insert sequence).
SOURCE	Homo sapiens testis CDNA to mRNA, clone_1fb:TEST12 clone:TEST12008847.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,

FEATURES	TITLE
	NEDO human cDNA sequencing project
	Unpublished
REFERENCE	2 (bases 1 to 1995)
AUTHORS	Isogai,T., Otsuki,T. and Sugiyama,T.
TITLE	Direct Submission
JOURNAL	Submitted (24-Oct-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory: 1532-3 Yana, Kisarazu, Chiba. 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3875, Fax:81-438-52-3886)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.
	Location/Qualifiers



```

source
1. .1955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TEST1200847"
/russne_type="testis"
/clone_id="TEST12"
/notes="cloning vector: pME18SFL3"
362. 757
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB71523.1"
/db_xref="GI:16553276"
/translation="MITIDLSIVIMP INRKVEFMSEVALPLSHGPKGSAVGSKSNK
PIGSESPSSSGRISCFVDELPTVIGLAVTAKTSKSESRPSRLRSQDAQDAACCF
SCVFVFPILISGCTKANKCOVVRKVI"

BASE COUNT      485 a      462 c      434 g      614 t
ORIGIN
Query Match      35.5%; Score 124.2; DB 9; Length 1995;
Best Local Similarity 58.7%; Pred. No. 8.6e+10;
Matches 192; Conservative 29; Mismatches 98; Indels 8; Gaps 3;

OY 1 ATGACCCTGCACTACCTGCACTTTACCCCACTGAGCCCTGCAAGTGTGGCCCGCTACCTTC 60
Db 861 ATGACCTACAGTGGCTGCTGCACTTTACCCCACTGAGCCCTGCAAGTGTGGCCCGCTACCTTC 920
OY 61 TGTACCTACGCGAGTTACCGCCCTGCTCTTCCTGCTCTGCTCTGCTCTTCACCTGACCCGCTT 120
Db 921 TGTACCTACGCGAGTTACCGCCCTGCTCTTCCTGCTCTGCTCTGCTCTTCACCTGACCCGCTT 977
OY 121 CACTGCATGCCAGAGGCGCCTGTATACAT -GCTTAGCANNNCACMCCTGCACNNNTGAGANA 179
Db 978 CACTGCATGCCAGAGTAATTGATTACCCATGCGCGAGGAGCTTACTGACTCTGCACACGACT 1037
OY 180 CTNCTNACCAATGATNATGATNGCTGCTGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 239
Db 1038 CTTTTCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1093
OY 240 GCAGACCAACNAAAGGACCTTNNAGACAGATGATTAANNACCNACTTACCTTAATNCT 299
Db 1094 AGATCCACACCTTCAGAGATTAATGATGCCCTCTTAATCAACGTGACACTGCTTACGACT 1153
OY 300 TTTCGAANNGAANAATCTTTCTTTGANA 326
Db 1154 GACGAGCAATCGCTGTGACACCAATGATTA 1180

RESULT 13
AF252826
LOCUS AF252826
DEFINITION Homo sapiens chromosome 8 clone CTD-237.N4 map 8p12, 8 unordered
pieces.
ACCESSION AF252826
VERSION AF252826.3 GI:14329036
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 126753)
AUTHORS Schillabel,M.-B., Baumgartl,C., Blechschlam,d.K., Detle,M., Jahn,N.,
Lehmann,R., Mendei,U., Polley,A., Reichwald,K., Schudy,A.,
Siddiqui,R., Taudien,S., Wen,G., Rosenthal,A. and Platzer,M.
Chromosome 8 genomic sequence
Unpublished
2 (bases 1 to 126753)
Genome Sequencing Center Jena.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT On Jun 8, 2001, this sequence version replaced gi:18151669.
-----
Center: Institute of Molecular Biotechnology

```

```

Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de

-----
Project Information
Center project name: H357
Center clone name: CYP-2373M4
-----
Summary Statistics
Sequencing vector: M13; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 121916 bases at least Q40
Consensus quality: 123958 bases at least Q30
Consensus quality: 125128 bases at least Q20
Quality coverage: 7.46 x in Q20 bases; sum-of-contigs

-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality 10.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
* 1 3657: contig of 3657 bp in length
* 3658 3757: gap of unknown length
* 3758 53867: contig of 50110 bp in length
* 53868 53967: gap of unknown length
* 53968 61529: contig of 7562 bp in length
* 61530 61629: gap of unknown length
* 61630 73303: contig of 11674 bp in length
* 73304 73403: gap of unknown length
* 73404 90798: contig of 17395 bp in length
* 90799 90898: gap of unknown length
* 90899 120417: contig of 29519 bp in length
* 120418 120517: gap of unknown length
* 120518 123961: contig of 3444 bp in length
* 123962 124061: gap of unknown length
* 124062 126753: contig of 2692 bp in length.
*
* Location/Qualifiers
*
* 1..126753
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="8"
* /map="8p12"
* /clone="CTD-2373M4"
*
BASE COUNT 37348 a 26850 c 26198 g 35657 t 700 others
ORIGIN
Query Match 35.5%; Score 124.2; DB 2; Length 126753;
Best Local Similarity 58.7%; Pred. No. 4.3e-09;
Matches 192; Conservative 29; Mismatches 98; Indels 8; Gaps
CY 1 ATGACCTGACGATTCCTGACCTTTACCCGACGATGAGCCGTCAGTGGGGCCGCTACCCG 60
DB 27548 ATGACCTGACGATTCCTGACCTTTACCCGACGATGAGCCGTCAGTGGGGCCGCTACCCG 27607
CY 61 TGTACCTGACGAGTAAAGGCGCTGCTTTTCACGCTGCGCCGCTTTCACCTGACCCGCTT 120
DB 27608 TGTACCTGACGAGTAAAGGCGCTGCTTTTCACGCTGCGCCGCTTTCACCTGACCCGCTT 27664
CY 121 CAGTCCATGCCAGAGGCGCTGTATACAT-GCCTAGGNNNNNCACMCTGCACNNNTGGANA 179
DB 27665 CAGTCCATGCCAGGTAATTTATACCATGAGCGAGGAGATTCGATCAGTACATGACCT 27724

```

```

TITLE
JOURNAL
COMMENT
Seaman, S., Severe, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Submitted (27-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 27, 2002 this sequence version replaced gi:17061055.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L20303
Center clone name: 2373_N4
-----
FEATURES
source
Location/Organisms
1. 127677
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="CTD-2373N4"
/clone_lib="CTD1 Human BAC"
61. 81
/rpt_family="AT_rich"
135. 285
/rpt_family="FAM"
290. 318
/rpl_family="CA)n"
578. 612
/rpl_family="CA)n"
/rpt_family="(TG)n"
complement(1567. 1751)
/rpt_family="MLTLEI"
complement(1798. 2049)
/rpt_family="Alusx"
2062. 2099
/rpt_family="(T)n"
complement(2759. 2844)
/rpt_family="LIMC5"
complement(2845. 3164)
/rpl_family="AluY"
complement(3165. 3390)
/rpt_family="LIMC5"
complement(3577. 3693)
/rpl_family="LIMC5"
4068. 4133
/rpt_family="FLAM_A"
complement(4959. 5211)
/rpl_family="LIMB2"
5207. 5245
/rpt_family="LIMB1"
5277. 5523
/rpl_family="L1MA8"
5529. 5660
/rpl_family="FLAM_A"
5662. 5696
/rpl_family="AluB"
5697. 5725
/rpl_family="(CAA)n"
5726. 5883
/rpl_family="AluB"
5902. 5951
/rpt_family="L1MA8"
5960. 5982
/rpl_family="AT_rich"
5991. 6138
/rpl_family="L1MA10"

```



TITLE  
JOURNAL  
COMMENT

Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, E., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Sever, P., Spencer, B., Stange-Thoman, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, C., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (26-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Nov 22, 2001 this sequence version replaced gi:16905307.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Project Information  
Center project name: 11917  
Center clone name: 115\_P4

# FEATURES

## SOURCE

1. 155926

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="RP11-115P4"

/clone\_lib="RP11-115P4 Human Male BAC"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

/rpt\_family="MIR"

```
repeat_region complement(26434..26589)
                /rpt_family="L1MC/D"
repeat_region complement(26837..26984)
                /rpt_family="L1MC/D"
repeat_region complement(26985..27115)
                /rpt_family="FLAM_A"
repeat_region complement(27117..27298)
                /rpt_family="L1MC/D"
repeat_region 27307..27329
                /rpt_family="(TAA)n"
```

Query Match 35.58; Score 124.2; DB 9; Length 153526;

Best Local Similarity 58.78; Pred. No. 4.6e-09;

Matches 192; Conservative 29; Mismatches 98; Indels 8; Gaps 3;

```
QY 1 ATGAGCTCACAGTACCTGCCTTTACCCGAGAGCCCTGAAAGTGGGCGCCGCTACCCCTG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 153239 ATGAGCTCACAGTACCTGCCTTTACCCGAGAGCCCTGAAAGTGGGCGCCGCTACCCCTG 153298

QY 61 TGTACCCAGCGAGATTAGCGCCCTGCTCTTCCTCCTCCGCTTCACCTACCGCGCTT 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 153299 TGTACCCAGCGAGATTAGCGCCCTGCTCTTCACCTCCTCCT ---GCTTGACCTATCCGCTT 153355

QY 121 CACTGCATGCCCGAGAGCGCCTGTATACAT GCCTAGGNNNNACAGCCTGACGNNNTGGANA 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 153356 CACTGCATGCCCGAGATTATGTATACCATCGGCGAGACTTTCACCTGACCGACAGAT 153415

QY 180 CTNCTNCGACACATGATNATGATNCGTNCGTGGGAATGNCCTGNCAGATGATCAAGCTACT 239
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 153416 CCTTTCAAACTTGGTTCGCCAGAGCACAC ---CATACAACTAACAGCTATACGACGTAC 153471

QY 240 GCAGACCAACNAGAGGAGCTTNGAGACAGAAATGATATNNAACNAGACTTATNCT 299
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 153472 AGATCCACCCTCAAGAGATTAAATGATCCCTCTAAATCAAGCTGACACCTCTTCAGGACT 153531

QY 300 TTGCAANCAANACTNTTCTTGGANA 326
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 153532 GACGAGATGCTGTGACAGCAATGATA 153558
```

Search completed: May 9, 2003, 20:19:09  
Job time : 1525 secs



GenCore version 5.1.4-p3.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 18:29:44 : Search time 219 Seconds  
(without alignments)  
3599.084 Million cell updates/sec

Title: US-09-880-711-328  
Perfect score: 350  
Sequence: 1 atgagccacagctactgca.....ttgaagattcacaataaa (350)

Scoring table: IDENTITY\_MUCDX  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125959159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

```
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50.4	25.8	845	21	AAC95479
2	65.2	28.6	632	24	ABN73285
3	65.2	28.6	632	24	ABN73375
4	59.4	27.0	3753	22	AAL04936
5	59.4	27.0	3753	23	ABL7830
6	58	26.6	617	23	ABV54245
7	58	26.6	16688	24	ABL32320
8	57.8	26.5	1782	22	AAF23414
9	57	26.3	8677	22	AAL36957

C	10	57	16.3	8680	22	AAL36955
C	11	56.6	16.2	7171	22	AAC88259
C	12	56.6	16.2	165199	24	ABK83460
C	13	56.4	16.1	820	24	ABN89955
C	14	56.4	16.1	831	21	AAC51574
C	15	56.4	16.1	843	21	AAC51564
C	16	56.4	16.1	1765	22	AAO52401
C	17	56.4	16.1	9060	12	AAO15132
C	18	56.4	16.1	9060	19	AAV15792
C	19	56.4	16.1	9060	22	AAO4542
C	20	56.4	16.1	9060	22	AAF23619
C	21	55.8	15.9	31952	22	AAK89370
C	22	55.6	15.9	3515	20	AAK60106
C	23	55.6	15.9	4874	20	AAK60103
C	24	55.4	15.8	2544	23	ABL21414
C	25	55.4	15.8	3454	23	ABL09688
C	26	55.4	15.8	5759	23	ABL22540
C	27	55.2	15.8	12400	23	ABL07678
C	28	55.2	15.8	13001	22	AAK82975
C	29	55.2	15.8	32185	22	AAI36806
C	30	55	15.7	3409	23	AAK75796
C	31	55	15.7	40267	24	AAK18958
C	32	55	15.7	159400	24	ABQ88125
C	33	54.8	15.7	396	22	AAI87801
C	34	54.8	15.7	621	21	AAC51567
C	35	54.8	15.7	777	21	AAC13950
C	36	54.8	15.7	845	21	AAC35459
C	37	54.8	15.7	879	21	AAC52193
C	38	54.8	15.7	954	20	AAK30342
C	39	54.8	15.7	1138	24	AAK562431
C	40	54.8	15.7	1496	24	ABK89681
C	41	54.8	15.7	1976	24	AAK62289
C	42	54.8	15.7	2035	19	AAV30292
C	43	54.8	15.7	2119	24	AAO31104
C	44	54.8	15.7	2720	22	AAH18544
C	45	54.8	15.7	85680	21	AAF22299

#### ALIGNMENTS

RESULT 1		
ID	AAC95479	standard; cDNA: 845 BP.
XX	AAC95479;	
XX	21-FEB-2001	(first entry)
XX	Human secreted protein gene 18 seq	ID NO:28.
KW	Human: secreted protein; immunosuppressive; antiarthritic; vulnary;	
KW	anti-rheumatic; antiproliferative; cycostatic; cardiac; vasotropic;	
KW	cerebroprotective; neuroprotective; antibacterial; virulide;	
KW	lignicide; ophthalmological; hyperproliferative disorder; neoplasm;	
KW	autoimmune disease; rheumatoid arthritis; cardiovascular disorder;	
KW	cerebrovascular disorder; cerebral ischemia; angiogenesis; viral;	
KW	nervous system disorder; Alzheimer's disease; bacteria; infection;	
KW	cell proliferation; skin aging; wound healing; chemotaxis;	
KW	food additive; ss.	
XX		
OS	Homo sapiens.	
XX		
XX	WO200058334-A1.	
XX		
PD	05-OCT-2000.	
XX		
PF	22-MAR-2000; 2000WO-US07507.	
XX		
PR	26-MAR-1999; 99US-0126594.	
PR	17-DEC-1999; 99US-0172408.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	

Human musculoskele  
African Green Monk  
Human cDNA differe  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Human secreted pro  
Zucchini ACC synth  
Zucchini ACC synth  
Zucchini CP-ACC 1A  
Zucchini ACC synth  
Human digestive sy  
Arabidopsis thalia  
Arabidopsis thalia  
Prosopilla melanoq  
Prosopilla melanoq  
Prosopilla melanoq  
Prosopilla melanoq  
Human immune/haema  
Human musculoskele  
DNA encoding novel  
Human cadherin-11k  
Human osteoblast g  
Human polynucleoti  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
DNA encoding a hum  
cDNA sequence #218  
Human polynucleoti  
cDNA sequence #76  
Bacillus thuringie  
Human Transcription  
Human cDNA sequenc  
BAC containing rep

PI Rosen CA, Ruben SM, Komatsoulis G;  
XX WPI: 2000-611701/58.  
DR P-PSDB: AAB51946.  
XX  
XX New nucleic acid molecules encoding 50 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -  
XX  
PS Claim 1; Page 338; 402pp; English.  
XX  
XX Polynucleotide sequences AAC95462 - AAC95511 represent cDNA encoding 50  
CC human secreted proteins AAB51929 - AAB51978. Sequences AAB51979 -  
CC AAB52010 represent alternative polypeptides encoded by the genes, and  
CC amino acid sequences with which they share homology. The genes and  
CC proteins have activities dependent on the tissues and cells in which they  
CC are expressed. Examples of their activities include immunosuppressive;  
CC antihistolic; antihemmatic; antiproliferative; cytostatic; cardiatic;  
CC vasotrophic; cerebroprotective; nootropic; neuroprotective; antibacterial;  
CC vincidine; fungicide; ophthalmological; and vulnary. The secreted  
CC proteins, polypeptides, antagonists and agonists may be useful in  
CC treating, preventing and/or diagnosing diseases and disorders such as  
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
CC e.g. cardiac arrest, cerebrovascular disorders e.g. Alzheimer's disease,  
CC anglogenesis, nervous system disorders e.g. Alzheimer's disease,  
CC infections caused by bacteria, viruses and fungi and ocular disorders  
CC e.g. corneal infection. The polypeptides can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamins, minerals, cofactors and other nutritional  
CC components. Oligonucleotides AAC95453 - AAC95461 and peptide AAB51928 are  
CC used in the isolation and characterisation of the proteins and  
CC polynucleotides of the invention.  
XX  
XX Sequence 845 BP; 226 A; 167 C; 182 G; 268 T; 2 other;  
SQ  
XX  
XX Query Match 25.8%; Score 90.4; DB 21; Length 845;  
Best local Similarity 56.8%; Pred. No. 3.9e-05;  
Matches 125; Conservative 29; Mismatches 56; Indels 10; Gaps 2  
QY 131 CCAGAGGCGCTGATACATGCGCTAGCGNNNNACNCTGCAGNNNTGGANCTNCTNACGA 150  
DB 605 CTAGTCTTTCTCTCCCTGAGGCAAGGCCAACACACCTGTTTGTCCACTTCTCCAGCA 654  
QY 191 CATGNTGATNGGCTGCTGCGGAATNGGCTGCACAGATCAACTACTGACAGCANCN 250  
DB 665 AATTACATTT--GTCCTCGGAATGTGTTTGAACATACCACTACTGACAGCCAGCA 721  
QY 251 AAGGAGCCTTNNAGACAGAGATGTATTANNAACCNACTACCTTATNCTTTGCCAANCA 310  
DB 722 GAGGAGCGCTCCCATGTGGAATTTGTTTGAAGCTA-----TTTCCGCCCTTTACAA 774  
QY 311 NAACNTTCTTTGANAATNTTGGAGAGATTGCATATAAA 350  
DB 775 AAACATATTTCTTGACGACCTTGAGAGATTTCATATAAAA 814  
RESULT 2  
ABN733285/  
ID ABN73285 standard; cDNA: 632 BP.  
XX ABN73285;  
XX  
XX 03-JUL-2002 (first entry)  
XX  
XX Bovine embryonic germ (EG) cell cDNA EST 000203a CONTIG 59.  
DE  
XX Bovine; Bos taurus; EST; expressed sequence tag; totipotence;  
KW

KW		development; gene; ss.
XX	Bos taurus.	
OS		
XX		
FN	WO200194550-A2;	
XX		
PD	13-DEC-2001.	
XX		
PF	07-JUN-2001; 2001WO-US18576.	
XX		
PR	07-JUN-2000; 2000US-209874P.	
XX	06-JUN-2001; 2001US-0876143.	
PA	(INF1-) INFILIN INC.	
XX		
PI	Eilertsen KJ, Pfister-Genskow M, Childs L;	
XX		
DR	WPI: 2002-351289/38.	
XX		
PT	An expressed sequence tag (EST), the expression of which, or its	
PT	complementary sequence, in a cell identifies the cell as a	
PT	developmentally competent or incompetent cell	
PS	Example 16; Page 142-143; 584pp; English.	
CC	The present invention describes an expressed sequence tag (EST), where	
CC	the EST is an isolated, enriched, or purified nucleic acid sequence	
CC	representing all or part of a gene, the expression of which, or its	
CC	complementary sequence, in a cell identifies the cell as a	
CC	developmentally competent or incompetent cell. Molecules which induce	
CC	tortipotence in one or more cells. Molecules which induce developmental	
CC	incompetence in a cell line are useful for preventing a full term	
CC	pregnancy in an animal and inhibiting tortipotence. The molecules are	
CC	also useful for treating a disease in an animal by inducing development	
CC	of one or more cells of the animal into a specific cell type. The	
CC	present sequence represents a bovine EST which is given in the	
CC	exemplification of the present invention.	
XS		
SQ	Sequence 632 BP; 191 A; 124 C; 152 G; 161 T; 4 other:	
	Query Match            18.6%; Score 65.2; DB 24; Length 632;	
	Best Local Similarity   48.2%; Pred. No. 0.85;	
	Matches   105; Conservative   29; Mismatches   63; Indels   21; Gaps   1;	
OY	154 AGNNNACNCCTGTGACNNNTGGANACTGTCAGCATGATGCATGCTCCTCGTGA 213      :::  ::  ::::  : ::  : ::  : ::  : ::  : ::  : ::	
DH	297 AGGCCAAGCCCCACACACCCTGTTTGTCCTTCCTCCACCAATAAGTTGCTCGGA 238   :::::  ::  ::::  : ::  : ::  : ::  : ::  : ::  : ::	
OY	214 ATNGCCCTGMCACAGATCAACCTACTGACGACCANCNAAGGAGCTTNNGAGACAGATG 273   :: ::  ::   ::   ::   ::   ::   ::   ::   ::   ::   ::	
DH	237 ATGTGTTTGAACATATCAACCTACTGACGACGAGGAGGAGCTCCCATGTAATT 178   :: ::  ::   ::   ::   ::   ::   ::   ::   ::   ::   ::	
OY	274 TATTANNAAACNACTACCTA-----TNCCTTGGCAANGANA 312   :: ::  ::   ::   ::   ::   ::   ::   ::   ::   ::   ::	
DH	177 GCCTTGACCTGTTTTTGTGTGTTGTTGTTGTTTTCCTTCCTCCCTTTTGCAAA 118   :: ::  ::   ::   ::   ::   ::   ::   ::   ::   ::   ::	
OY	313 ACTNTCTTTGANMAATNTTGCAGATTTGCATATAAAA 350   :: ::  ::   ::   ::   ::   ::   ::   ::   ::   ::   ::	
DH	117 ACTATTCTTGTAACGTTTGAGAGATTTCAATAAAAA 80   :: ::  ::   ::   ::   ::   ::   ::   ::   ::   ::   ::	
RESULT 3		
ID	ABN73375/c	
XX	- ABN73375 standard; CDNA; 632 BP.	
XX		
AC	ABN73375;	
XX		
DT	03-JUL-2002 (first entry)	
XX		
DE	Bovine embryonic germ (EG) cell cDNA EST 000203a CONTIG 59.	
XX		
XX	Bovine; Bos taurus; EST; expressed sequence tag; totipotency;	



KW development: gene; ss.  
XX  
OS Bos taurus.  
XX  
PN W0200194550-A2.  
XX  
PD 13-DEC-2001.  
XX  
PF 07-JUN-2001: 2001W0-US18575.  
XX  
PR 07-JUN-2000: 2000US-209874P.  
XX  
PR 06-JUN-2001: 2001US-0876143.  
XX  
PA (INF1-) INFGEN INC.  
XX  
PI Ellertsen KJ, Pfister-Genskow M, Childs L;  
XX  
DR WPL: 2002-351289/38.  
XX  
PT An expressed sequence tag (EST), the expressor of which, or its  
PT complementarily sequence, in a cell identifies the cell as a  
PT developmentally competent or incompetent cell.  
XX  
PS Example 16: Page 160; 584pp; English.  
XX  
CC The present invention describes an expressed sequence tag (EST), where  
CC the EST is an isolated, enriched, or purified nucleic acid sequence  
CC representing all or part of a gene, the expression of which, or its  
CC complementarily sequence, in a cell identifies the cell as a  
CC developmentally competent or incompetent cell. Molecules which induce  
CC developmental competence in a cell line are useful for inducing  
CC totipotency in one or more cells. Molecules which induce developmental  
CC incompetence in a cell line are useful for preventing a full term  
CC pregnancy in an animal and inhibiting totipotency. The molecules are  
CC also useful for treating a disease in an animal by inducing development  
CC of one or more cells of the animal into a specific cell type. The  
CC present sequence represents a bovine EST which is given in the  
CC exemplification of the present invention.  
XX  
SO Sequence 632 BP; 191 A; 124 C; 152 G; 161 T; 4 other:  
  
Query Match 18.6%; Score 65.2; DB 26; Length 632;  
Best Local Similarity 48.2%; Pred. No. 0.85;  
Matches 105; Conservative 29; Mismatches 63; Indels 21; Gaps 1;  
  
QY 154 ACCNNNNACNCCTGACNNTGGANACTGTCAGACATGATGATGTCGTCGGA 213  
DB 297 AGCGAAGCCGACACACCTGTTTTCCTTCCTTCGACCAATAGATTGTCCTGGGA 238  
QY 214 ATGNGCCGNCACAGATCACTACTGACAGACCAGCAGAGTCTTNGACAGCAATG 273  
DB 237 ATGTGTTTGAAGATTCAGCTACTGACAGACGACAGAGGAGTCCATGTTGAAATT 178  
QY 274 TATTANMAACNACACCTTA-----TNCCTTTTGCAGAACANA 312  
DB 177 GCCTGTACCTGTTTCTGCTGCTGTTGTTTTCCTTCCTTCCTTTTGGAAA 118  
QY 313 ACTNTCTTGGANAATNTTGGAGATTCATATAAA 350  
DB 117 ACTATTCTTGCTACGTTTGAGACATTCATAAATAA 80  
  
RESULT 4  
AAL04936/C  
ID AAL04936 standard; DNA: 3753 BP.  
XX  
AC AAL04936;  
XX  
PT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen DNA seq ID NO: 7624.  
XX  
KW Human: reproductive system related antigen; reproductive system disorder;

KW cancer: gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN W0200155320-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001: 2001W0-US01339.  
XX  
PR 31-JAN-2000: 2000US-0179065.  
XX  
PR 04-FEB-2000: 2000US-0180628.  
XX  
PR 24-FEB-2000: 2000US-0184664.  
XX  
PR 02-MAR-2000: 2000US-0186350.  
XX  
PR 16-MAR-2000: 2000US-0189874.  
XX  
PR 17-MAR-2000: 2000US-0190076.  
XX  
PR 18-APR-2000: 2000US-0198123.  
XX  
PR 19-MAY-2000: 2000US-0205515.  
XX  
PR 07-JUN-2000: 2000US-0209467.  
XX  
PR 28-JUN-2000: 2000US-0214886.  
XX  
PR 30-JUN-2000: 2000US-0215135.  
XX  
PR 07-JUL-2000: 2000US-0216647.  
XX  
PR 11-JUL-2000: 2000US-0216880.  
XX  
PR 11-JUL-2000: 2000US-0217487.  
XX  
PR 14-JUL-2000: 2000US-0217496.  
XX  
PR 26-JUL-2000: 2000US-0218290.  
XX  
PR 26-JUL-2000: 2000US-0220963.  
XX  
PR 26-JUL-2000: 2000US-0220964.  
XX  
PR 14-AUG-2000: 2000US-0224518.  
XX  
PR 14-AUG-2000: 2000US-0224519.  
XX  
PR 14-AUG-2000: 2000US-0225213.  
XX  
PR 14-AUG-2000: 2000US-0225214.  
XX  
PR 14-AUG-2000: 2000US-0225266.  
XX  
PR 14-AUG-2000: 2000US-0225267.  
XX  
PR 14-AUG-2000: 2000US-0225268.  
XX  
PR 14-AUG-2000: 2000US-0225270.  
XX  
PR 14-AUG-2000: 2000US-0225447.  
XX  
PR 14-AUG-2000: 2000US-0225757.  
XX  
PR 14-AUG-2000: 2000US-0225758.  
XX  
PR 14-AUG-2000: 2000US-0225759.  
XX  
PR 18-AUG-2000: 2000US-0226279.  
XX  
PR 22-AUG-2000: 2000US-0226681.  
XX  
PR 22-AUG-2000: 2000US-0226868.  
XX  
PR 22-AUG-2000: 2000US-0227182.  
XX  
PR 23-AUG-2000: 2000US-0227009.  
XX  
PR 30-AUG-2000: 2000US-0228924.  
XX  
PR 01-SEP-2000: 2000US-0229287.  
XX  
PR 01-SEP-2000: 2000US-0229343.  
XX  
PR 01-SEP-2000: 2000US-0229344.  
XX  
PR 01-SEP-2000: 2000US-0229345.  
XX  
PR 05-SEP-2000: 2000US-0229509.  
XX  
PR 05-SEP-2000: 2000US-0229513.  
XX  
PR 06-SEP-2000: 2000US-0230437.  
XX  
PR 06-SEP-2000: 2000US-0230438.  
XX  
PR 08-SEP-2000: 2000US-0231242.  
XX  
PR 08-SEP-2000: 2000US-0231243.  
XX  
PR 08-SEP-2000: 2000US-0231244.  
XX  
PR 08-SEP-2000: 2000US-0231413.  
XX  
PR 08-SEP-2000: 2000US-0231414.  
XX  
PR 08-SEP-2000: 2000US-0231414.  
XX  
PR 08-SEP-2000: 2000US-0232080.  
XX  
PR 12-SEP-2000: 2000US-0232081.  
XX  
PR 12-SEP-2000: 2000US-0231568.  
XX  
PR 14-SEP-2000: 2000US-0232397.  
XX  
PR 14-SEP-2000: 2000US-0232398.  
XX  
PR 14-SEP-2000: 2000US-0232399.  
XX  
PR 14-SEP-2000: 2000US-0232399.  
XX  
PR 14-SEP-2000: 2000US-0232400.  
XX  
PR 14-SEP-2000: 2000US-0232401.  
XX  
PR 14-SEP-2000: 2000US-0233063.  
XX  
PR 14-SEP-2000: 2000US-0233064.  
XX  
PR 14-SEP-2000: 2000US-0233065.  
XX  
PR 21-SEP-2000: 2000US-0234223.  
XX  
PR 21-SEP-2000: 2000US-0234274.  
XX  
PR 25-SEP-2000: 2000US-0234597.

XX	PA	(HUMA-)	HUMAN GENOME SCI INC.
XX	PI	Rosen CA,	Barash SC, Ruben SM;
XX	DR	WPI:	2001-465570/50.
XX	PT	Isolated nucleic acid molecule encoding a reproductive system antigen	
XX	PP	is used in preventing, treating or ameliorating a medical condition	
XX	PS	Disclosure; SEQ ID NO 7624; 1297pp + Sequence Listing; English.	
XX	CC	The present invention provides the protein and coding sequences of a	
XX	CC	number of human reproductive system related antigens. These can be used	
XX	CC	in the prevention and treatment of reproductive system disorders,	
XX	CC	including cancer. The present sequence is a genomic sequence encoding	
XX	CC	protein of the invention.	
SO	Sequence	3753 BP; 1058 A; 746 C; 783 G; 1166 T; 0 other;	
	Query Match	17.0%; Score 59.4; DB 22; Length 3753;	
	Best Local Similarity	40.3%; Pred. No. 79;	
	Matches 102; Conservative	30; Mismatches 121; Indels 0; Gaps	
OY	80	GCGTCCTCTTCCNCCCTCCGCGCTTTCAACACTAACCCCGCTGCAGTGATGCCACAGGCN	139
Db	3011	GCCAGGATGTTCATCATCTCCTACCTCGTAGTCCGCCCATGTGGCTCCCAAAATGCT	2952
OY	140	CTGTATACAGCCCTAGGNNNCANCCCTGACNNNTGGANAACCTCTNCGACACATGNATGA	199
Db	2951	GGGATTTCAGGCGTAAGCCACCATCTGCGCGCTTACCAAAATATTTTCAACACTATTCCA	2892
OY	200	TNCGTCNCGGGAATANGCCCTCNACACATCACTACTGCAGACCANCNAAGGAGCT	259
Db	2891	TCITTTATTTTCTACTGCACAACCTGGAGAGGTAGGACGTACGGCAGATTATTATTTGCA	2832
OY	260	TNNCAGACGAATGTATTANNACCNACTACCTTATNCTTTTGGAAANGAANAACCTTTC	319
Db	2831	TTTTATCAAGATTGTATATATAATATACCATTTTTCGATGTCTGGAAGTAACTTGGCT	2772
OY	320	TTTGANAATMTT 332	
Db	2771	TTGCATGATCTAT 2759	
ID	ABL97830/c	Standard; DNA; 3753 BP.	
XX	ABL97830;		
XX	21-JUN-2002	(first entry)	
DE	Human testicular antigen encoding DNA fragment SEQ ID NO: 2482.		
KW	Human; testicular antigen; testes; cancer; metastasis; immune disorder		
KM	reproductive system disorder; urinary system disorder; gene therapy;		
KM	cardiovascular disorder; respiratory disorder; neurological disorder;		
KM	gastrointestinal disease; infection; cytostatic; gene; ds.		
OS	Homo sapiens.		
PN	WO200155317-A2.		
PD	02-AUG-2001.		
PE	17-JAN-2001; 2001WO-US01329.		
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		



cc The invention relates to an isolated nucleic acid molecule (I) compris

EA (EPIG-) EPIGENOMICS AG.

[illegible]

XX WP1: 2001-025020/03.  
DR P-PSDB: AAB49653.

XX  
PT New SPFX polypeptides and nucleic acids useful for treating or  
PT preventing cancer, other disorders related to angiogenesis,  
PT neurodegenerative diseases, autoimmune disorders and allergic reactions

XX  
PS Claim 9; Page 20-21; 1x2pp; English.

XX  
CC polynucleotide sequences AAF23410 - AAF23419 encode secreted SPFX  
CC proteins AAB95649 - AAB95658. Sequences AAF23420 - AAF23450 represent  
CC primers and probes used in the isolation and characterisation of the SPFX  
CC DNA sequences of the invention. The new polypeptides and nucleic acids  
CC can be used in screening assays, detection assays, preventive or  
CC predictive medicine, therapeutic and prophylactic treatment, and  
CC pharmacogenomics. Specifically, the SPFX polypeptides and nucleic acids  
CC are useful for treating cancer; other disorders related to angiogenesis  
CC e.g. abnormal wound healing, psoriasis; neurodegenerative diseases;  
CC immune disorders; liver cirrhosis; benign tumours; fibrotic conditions  
CC and tissue hypertrophy (e.g. benign prostatic hypertrophy); allergic  
CC reactions and conditions such as asthma and other respiratory problems;  
CC as well as in treating or preventing diseases associated with organ  
CC transplantation, atherosclerosis-associated diseases or disorders. The  
CC polypeptides can also be used for bone, cartilage, tendon, ligament  
CC and/or tissue growth or regeneration, wound healing, tissue repair and  
CC replacement, gut protection or regeneration, as a contraceptive, to  
CC inhibit tumours, infections caused by bacteria, virus, fungi and other  
CC parasites, and as a vaccine. SPFX antibodies may be used to isolate or  
CC detect SPFX proteins, monitor protein level in tissue as part of a  
CC clinical testing procedure, treat proliferative disorders including  
CC tumours and benign hyperplasias.

XX  
SO Sequence 1782 BP; 521 A; 416 C; 397 G; 448 T; 0 other:

Query Match: 16.5%; Score 57.8; DB 22; Length 1782;  
Best Local Similarity: 40.5%; Pred No. 62;  
Matches: 83; Conservative: 30; Mismatches: 92; Indels: 0; Gaps: 0

OY 139 NCTGATATATCCTCAGCGNNNCACNCTGCAGNNNTGGANACTCTTCAGCACATCATC 158  
          | | | | |     | : : : |     | : : | | : | : |  
DB 185 AGTGTATGCTCAATAAATATTCATTGACGTTTGTTGGTAGCTGATGCCGACGAATCGG 244  
OY 199 ATNCGCTCTTCGGCATGCGCTGACGACGATCAACCTATCTCCAGACGANCNAAGGAGC 258  
          | | | | |     | : : | | | | | | | | | | | |  
DB 245 ATGCTCAATCTGGAAGCTAGCTGACATCCCATCAAGAAGAAAAACAGACAGCTGACCA 304  
OY 259 TTNNAGAGTACGATATATPANNAAACNACTACTTATCTCTTGGCAAGNAAATNTT 318  
          | : | | | | | : : | | | | | | | | | | | | | | |  
DB 305 CTACATGACAGAAAGCTACAGACGCAATTCATATATGAGTTGTAATTGCAATTACATCT 364  
OY 319 CTTTGGANAATTTTTGAGACATTTCGA 343  
          | | | : | | | | | | | | |  
DB 365 CTTTAAAATATATGTGCCACACTTTCGA 389

RESULT 9  
AAL36957/c  
ID AAL36957 standard: DNA: 8677 BP.

XX  
XX AAL36957:  
AC  
DE Human musculoskeletal system related polynucleotide SEQ ID NO 3322.  
XX  
XX 08-JAN-2002 (first entry)  
XX  
XX  
KW Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;  
KW anti-allergic; hepatocentric; antidiabetic; anti-inflammatory; anticancer;  
KW cardiovascular; anticoagulant; antibacterial; antifungal; antiparasitic;  
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ds





XX	Rosen CA, Barash SC, Ruben SM;
PI	WPI. 2001-451937/48.
DR	
XX	Isolated polypeptide for treating, preventing and/or prognosing
PI	disorders related to the musculoskeletal system including
PI	musculoskeletal cancers and also for testing and detection e.g.
PI	diagnosis -
XX	
PS	Example 2: SEQ ID NO 3321; 781bp + Sequence Listing; English.
XX	
CC	The invention relates to novel genes (AAU34669-AAU37666) and proteins
CC	(AB03087-AB04109) associated with the musculoskeletal system useful
CC	for preventing, treating or ameliorating medical conditions e.g. by
CC	protein or gene therapy. The genes are isolated from a range of human
CC	tissues disclosed in the specification. The nucleic acids, proteins,
CC	antibodies and (ant)agonists are useful in the diagnosis, treatment
CC	and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC	other cancers of the adrenal gland, bone, bone marrow, breast,
CC	gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC	anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC	anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC	(c) cardiovascular disorders such as myocardial ischaemia; (d) wound
CC	healing; (e) neurological diseases such as viral, bacterial, fungal and
CC	and (f) infectious diseases such as viral, bacterial, fungal and
CC	parasitic infections.
CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX	
SQ	Sequence 8680 BP; 2010 A: 2541 C; 2538 G; 1591 T; 0 other;
	Query Match 16.3%; Score 57; DB 22; Length 8680;
	Best Local Similarity 40.2%; Pred. No. 5.8e+02;
	Matches 84; Conservative 30; Mismatches 95; Indels 0; Gaps 0;
QY	121 CACTCATGATGCCAGAGGCGNCCTGTATACATGCTCTAGCANNNNCCAGCNCCTGACNNNTGGANAC 180
Db	4979 CACTTACGGGCTTCACAGAGGCTGTATACATCACCCCAATGAGTAGACTTCTGGGGTC 4920
QY	181 TNCCTNACACATGATATATNCGTCNCTGGCAATNCGCTGNCACAGATCAACTACTG 240
Db	4919 AACTGATCGGCGCTGTGCTCCCTGCGCAGGCGTGTGACACTGCAGACAAAGCCATTTCTCCNC 4860
QY	241 CAGACCACNMAAGGAGGCTTNNAGACAGAAATGTATTNNAAACNACATCACTTATNCCT 300
Db	4859 AGGACCCCAAGGGGAGCTGCAGGGCCAGAAATGTCTTCAACGCTCGTACTCTGCGAGAC 4800
QY	301 TTGCAAAANNAACATNTTCTTTGANAATN 329
Db	4799 TGACTGCTGCGCCGCTTGGCTTGATG 4771
RESULT 11	
AC	AAC88299/c
ID	AAC88299 standard; DNA: 7171 BP.
XX	
AC	AAC88299;
XX	
DT	02-MAR-2001 (first entry)
XX	
DE	African Green Monkey simian hepatitis A viral isolate AGM-27.
XX	
XX	African green monkey; hepatitis A; vaccine; attenuated strain;
KW	chimeric virus; ds.
XX	
OS	Simian hepatitis A virus.
PN	US6146643-A.
XX	
14-NOV-2000	



```
XX 07-JUN-1995; 9505-0473185.
XX
XX 26-MAR-1991; 9105-0674852.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Purcell RH, Tsarev SA, Emerson SU;
XX
XX WPI: 2001-040424/05.
XX
XX Novel human/simian chimeric hepatitis A virus for use as vaccine
XX against hepatitis A, comprises human attenuated hepatitis A virus
XX sequence whose specific regions are replaced by simian hepatitis A
XX virus AGM-27.
XX
XX Disclosure: Column 7-14; 14pp: English.
XX
XX The present invention provides a chimeric hepatitis A viral genome
XX comprising the immunogenic properties of the human hepatitis A virus and
XX the nonstructural regions of the simian hepatitis A genome. This can be
XX used to produce an attenuated strain which has a use as a vaccine against
XX hepatitis A in humans.
XX
XX Sequence 7171 BP; 2058 A; 1171 C; 1599 G; 2343 T; 0 other:
XX
XX Query Match 16.2%; Score 56.6; DB 22; Length 7171;
XX Best local Similarity 40.1%; Pred. No. 5.5e+02;
XX Matches 83; Conservative 30; Mismatches 94; Indels 0; Gaps 0;
XX
XX 135 NCTGATACATCCCTAGAGNNNNACGCTGACNNNTGAGNACTNCTNCAGCATGATG 198
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 5412 CATCAATTCACAAAGCTCTTGAAACATCATCTCTTAAATGCTGATTTATCTCTA 5353
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 199 ATNGGCTCNCCTGCGATGACCTGACAGATCACTACTGACAGCCAGCAGGAGAGC 258
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 5332 CTTTGGAATTTAGAGACCTTATTAANNAACCNACTACCTATNCTTTGCCAANCAAGTNTT 5293
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 259 TTNNCA3ACAGCATGATTAATTAANNAACCNACTACCTATNCTTTGCCAANCAAGTNTT 318
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 5292 TTACACACATTTACACAGAGAAATTTGAATTAATAGTTCTCTCTTTATTAATAAAGCATTC 5233
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX 319 CTTGAGANATNTTTGAGAGATTTTCAT 345
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 5232 ATTGATATATCTTCTCAAAATTTGTAT 5205
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX RESULT 12
XX ABR83460/c
XX ID ABR83460 standard: cDNA: 165199 BP.
XX
XX ABR83460;
XX
XX 14-AUG-2002 (first entry)
XX
XX Human cDNA differentially expressed in granulocytic cells #31.
XX
XX Human: ss; granulocytic cell; DNA chip; bacterial infection;
XX viral infection; parasitic infection; protozoal infection;
XX fungal infection; sterile inflammatory disease; psoriasis;
XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
XX cardiac reperfusion injury; renal reperfusion injury; AIDS;
XX adult respiratory distress syndrome; inflammatory bowel disease;
XX Crohn's disease; ulcerative colitis; periodontal disease;
XX granulocyte activation; chronic inflammation; allergy.
XX
XX Homo sapiens.
XX
XX WO200228999-A2.
XX
XX 11-Apr-2002.
XX
XX 03-OCT-2001; 2001WO-US30821.
```

```
XX 03-OCT-2000; 2000US-237189P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
XX WPI: 2002-435428/46.
XX
XX Detecting granulocyte activation by detecting differential expression
XX of genes associated with granulocyte activation, which serves as
XX diagnostic markers that is useful for monitoring disease states and
XX drug toxicity.
XX
XX Claim 1; SEQ ID No 31; 114pp: English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing
XX the expression level to an expression level in an unactivated
XX GC, where differential expression of Gs is indicative of GCA.
XX Also included are modulating (M2) GCA by contacting GC with an agent
XX that alters the expression of at least one gene in Gs; (2) screening (M3)
XX for an agent capable of modulating GCA or an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease using the
XX gene expression profile; (3) detecting (M4) an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease, by detecting the
XX level of expression in a sample of the tissue of gene(s) from Gs, where
XX the level of expression of the gene is indicative of inflammation;
XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,
XX an allergic response in a subject, exposure of a subject to a pathogen
XX or sterile inflammatory disease, by contacting a tissue having
XX inflammation with an agent that modulates the expression of gene(s)
XX from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
XX modulating GCA; M3 is useful for screening an agent capable of modulating
XX GCA preferably in an inflammation in a tissue; M4 is useful for
XX detecting an inflammation (especially chronic) in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease (e.g., psoriasis, rheumatoid arthritis,
XX glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
XX reperfusion injury, AIDS, adult respiratory distress syndrome,
XX inflammatory bowel disease, Crohn's disease, ulcerative colitis,
XX periodontal disease; also bacterial infection, viral infection,
XX parasitic infection, protozoal infection, fungal infection and M5 is
XX useful for treating one of the above conditions. The present
XX sequence represents a gene differentially expressed in granulocytes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from Wipo at
XX ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 165199 BP; 48510 A; 33223 C; 34406 G; 49060 T; 0 other:
XX
XX Query Match 16.2%; Score 56.6; DB 24; Length 165199;
XX Best local Similarity 40.8%; Pred. No. 1e+04;
XX Matches 78; Conservative 29; Mismatches 84; Indels 0; Gaps 0;
XX
XX 136 GACNCTGTATACATGCTAGAGNNNNACGCTGACNNNTGAGNACTNCTNCAGCATGATG 195
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 73582 GTGATATACATTAATGCTAAGAGGATCATGACAGCATGAGCTATTCCTAATAATATAT 73524
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX 196 ATGATNAGCTCNCCTGCGAATGACGCTGACAGATCACTACCTACCTGAGAGCCAGCAGG 255
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 73522 ATGACACCCATACAGAGAGCCGAGATCAATGAAGCTTCTTACGAGCATCAAGAAGAG 73453
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX 256 ACCTTNNAGACAGCATGATTAATTAANNAACCNACTACCTATNCTTTGCCAANCAAGTNTT 315
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 73462 ACTTAGATATACCATATGATAGTAAGTAGAGAGACTTTAATATCCAGTGCATATATTAAGCA 73403
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX 316 NTTCCTTTGANA 326
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
```

DB 73402 GATCAATGACA 73392

RESULT 13

ABN98965

ID ABN98965 standard; DNA; 820 BP.

XX

AC ABN98965;

XX

DT 01-AUG-2002 (first entry)

XX

DE Arabidopsis thaliana, expressed polynucleotide SEQ ID NO 733.

XX

KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress; disease; crop; thale cress; tolerance factor; insect; pathogen; nutrition; ds.

XX

OS Arabidopsis thaliana.

XX

PN US2002023281-A1.

XX

PD 21-FEB-2002.

XX

PF 26-JAN-2001; 2001US-0770445.

XX

PR 27-JAN-2000; 2000US-178472P.

XX

PA (GORL/) GORLACH J.

PA (ANY/) AN Y.

PA (HAMI/) HAMILTON C M.

PA (PRIC/) PRICE J L.

PA (RAIN/) RAINES T M.

PA (YUY/) YU Y.

PA (RAME/) RAMEKA J G.

PA (PAGE/) PAGE A.

PA (MATH/) MATHW A V.

PA (LEDF/) LEDFORD B L.

PA (WESS/) WESSNER J P.

PA (HAAS/) HAAS W D.

PA (GARC/) GARCIA C A.

PA (KRICK/) KRICKER M.

PA (SLAT/) SLATER T.

PA (DAVI/) DAVIS K R.

PA (ALEE/) ALLEN K.

PA (HOFF/) HOFFMAN N.

PA (HURB/) HURBAN P.

XX

PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y, Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD, Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N, Hurban P;

PI

DR WPI: 2002-400781/43.

XX

PT New Arabidopsis thaliana nucleic acid for identifying homologous genes, producing compositions that modulate the expression or function of its encoded protein, and mapping functional regions of protein

PT

PS Claim 1; SEQ ID NO 733; 49bp + Sequence Listing; English.

XX

CC The invention relates to an Arabidopsis thaliana nucleic acid (I) comprising a sequence capable of hybridizing under stringent conditions to a sequence selected from any one of 999 sequences (ABN98233-ABN99231), given in the specification or its fragment. A polypeptide (II) encoded by (I), a transgenic plant (III) comprising an exogenous nucleic acid for a genetically modified cell (IV) comprising an exogenous nucleic acid, is useful for screening a candidate agent for its biological effect. (I) is useful in identifying homologous or related genes, in producing compositions that modulate the expression or function of its encoded protein, mapping functional regions of the protein and in studying associated physiological pathways. (I) is also useful for the genetic manipulation of cells, particularly plant cells. (I) is also useful in screening assays of various plant strains to determine the strains that

CC are best capable of withstanding a particular disease or environmental stress. (II) and (III) are useful for screening of biologically active agents, e.g. fungicides, insecticides, etc., for elucidating biochemical pathways. The screened agents are useful in improved methods of treating crops to prevent or treat disease. (II) are also useful in screening programs to identify agents that mimic or enhance the action of tolerance factors. Such agents are useful in improved methods of treating crops to enhance their tolerance to environmental stress. (I) is also useful for enhancing or inhibiting production of a biosynthetic product in a plant. (III) is useful for identifying other mediators that may induce expression of proteins of interest, for establishing the extent to which any specific insect and/or pathogen is responsible for damage to a particular plant, for identifying other mediators that enhance or induce tolerance to environmental stress, for identifying factors involved in biosynthetic pathways of nutritional, commercial, or medicinal value and for identifying productions of nutritional, commercial or medicinal value. (IV) is useful in the study of genetic function and regulation, for alteration of the cellular metabolism and for screening compounds that may affect the biological function of the gene or gene products. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at [seqdata.uspto.gov/sequence.html?docID=999099770445](http://seqdata.uspto.gov/sequence.html?docID=999099770445).

CC Sequence 820 BP; 230 A; 185 C; 193 G; 212 T; 0 other;

Query Match 16.1%; Score 56.4; DB 24; Length 820;

Best Local Similarity 39.7%; Pred. No. 43;

Matches 96; Conservative 30; Mismatches 116; Indels 0; Gaps 0;

QY 88 TTCCTCCTCCCTCCGCTTACCTACCCCGCTTACCTGATGCCAGAGCNCCTATAC 147

DB 189 TCCCTGCTCCACTCTGTGACCTCCGCCCTGTGTGATGCTGACACTTAACTAGTACA 248

QY 148 ATGCCCTAGGNNNNCAACNCCCTGACNNNTGGANACTGNCAGACATGATGATNCCGTG 207

DB 249 ACATGAGGTGACGACACCGGTAAAGGATTGACCTTGAGAGGCTCAAGCTCGGTATGCC 308

QY 208 CTGGGATGNGCCTCNCACAGATCAACCTACCTACGACGACGACANNAAGGAGCTTNGAGAC 267

DB 309 CAAAGAGTGTGGCCGCTTCAATGTTGCTGTGTGACCATGCTGCAAGACCATCTT 368

QY 268 AGAATGTATTANNAACNCACTTATNCTTTTGGCAANGANNACTNTCTTTGGANNA 327

DB 369 TGGAGGGTCTTCAACCAATGTCAGAGGCTGAACCTACAGACCAAGTATGATTT 428

QY 328 TN 329

DB 429 TC 430

RESULT 14

AAC51574

ID AAC51574 standard; DNA; 831 BP.

AC AAC51574;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 69015.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR

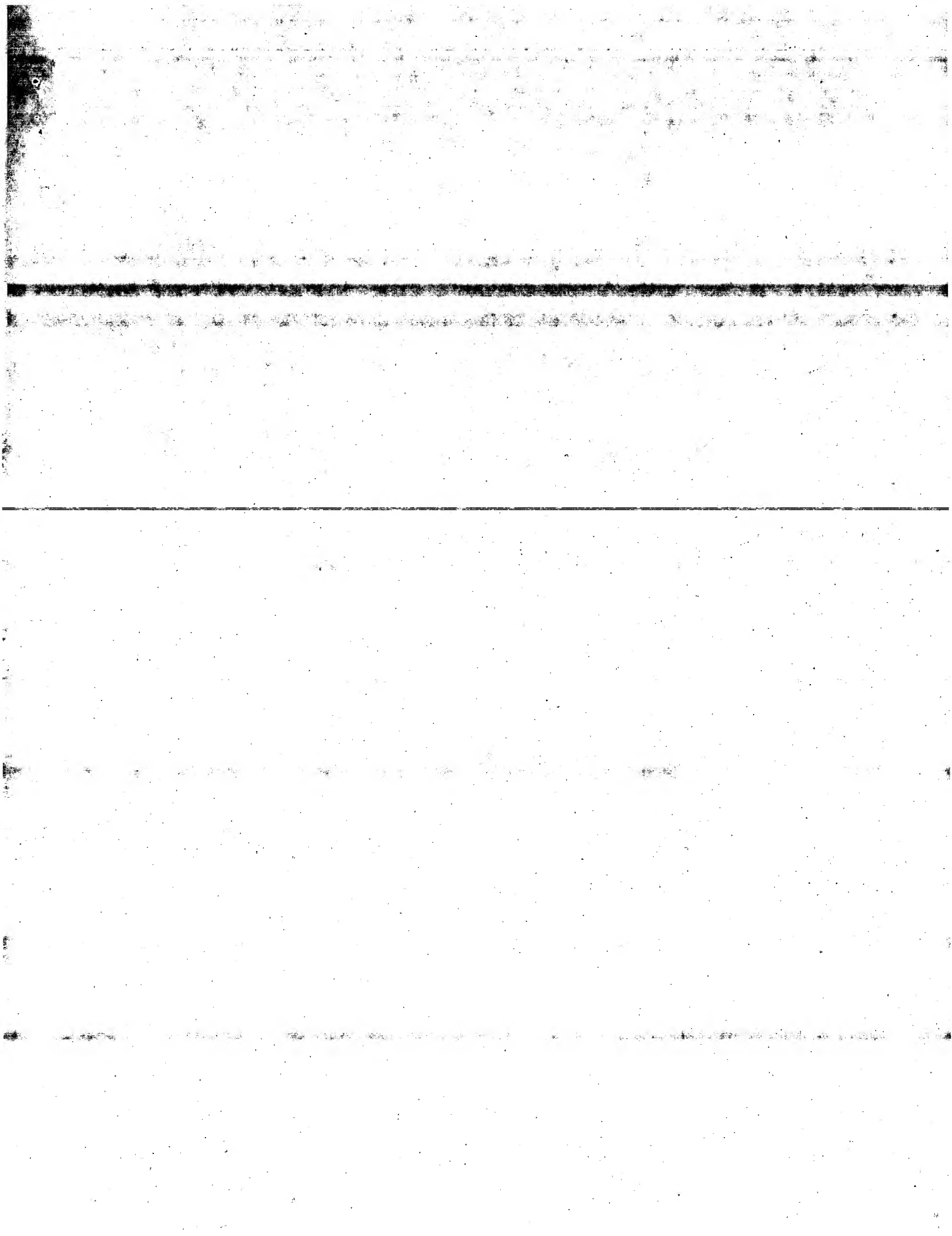
PR 05-MAR-1999; 990S-0123180.  
PR 05-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125758.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 23-APR-1999; 990S-0130891.  
PR 28-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 30-APR-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 06-MAY-1999; 990S-0132487.  
PR 07-MAY-1999; 990S-0132863.  
PR 11-MAY-1999; 990S-0134256.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139452.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139896.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.

PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145084.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145152.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146389.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146387.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147453.  
PR 09-AUG-1999; 990S-0147535.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149358.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149502.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.  
PR 26-AUG-1999; 990S-0150684.  
PR 27-AUG-1999; 990S-0151065.  
PR 27-AUG-1999; 990S-0151066.  
PR 27-AUG-1999; 990S-0151080.  
PR 30-AUG-1999; 990S-0151303.  
PR 31-AUG-1999; 990S-0151438.  
PR 01-SEP-1999; 990S-0151930.  
PR 07-SEP-1999; 990S-0152363.  
PR 10-SEP-1999; 990S-0153070.  
PR 13-SEP-1999; 990S-0153758.  
PR 15-SEP-1999; 990S-0154018.  
PR 16-SEP-1999; 990S-0154039.  
PR 20-SEP-1999; 990S-0154779.  
PR 22-SEP-1999; 990S-0155139.  
PR 23-SEP-1999; 990S-0155548.  
PR 24-SEP-1999; 990S-0155659.  
PR 28-SEP-1999; 990S-0156458.  
PR 29-SEP-1999; 990S-0156536.  
PR 04-OCT-1999; 990S-0157117.

DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 68379.
XX	
KX	
KW	Hybridisation assay; genetic mapping; gene expression control
KW	protein identification; signal transduction pathway;
KW	metabolic pathway; promoter; termination sequence; ss.

OS	Archidopsis thaliana.	
XX	EpI033405-A2.	
PN		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000BP-0301A439	
XX		
PR	25-FEB-1999.	9905-01218235
PR	05-MAY-1999.	9905-01231600
PR	03-MAY-1999.	9905-01253488
PR	23-MAY-1999.	9905-01257588
PR	25-MAY-1999.	9905-01262624
PR	29-MAY-1999.	9905-01262645
PR	01-APR-1999.	9905-01274632
PR	06-APR-1999.	9905-01282244
PR	08-APR-1999.	9905-01287714
PR	16-APR-1999.	9905-01298455
PR	19-APR-1999.	9905-01300787
PR	21-APR-1999.	9905-01304459
PR	23-APR-1999.	9905-01305010
PR	28-APR-1999.	9905-01310491
PR	30-APR-1999.	9905-01311489
PR	04-MAY-1999.	9905-01322048
PR	05-MAY-1999.	9905-01322407
PR	06-MAY-1999.	9905-01322444
PR	06-MAY-1999.	9905-01324465
PR	07-MAY-1999.	9905-01324876
PR	11-MAY-1999.	9905-01328653
PR	14-MAY-1999.	9905-01342856
PR	14-MAY-1999.	9905-01344219
PR	14-MAY-1999.	9905-01344211
PR	18-MAY-1999.	9905-01343731
PR	19-MAY-1999.	9905-01347461
PR	20-MAY-1999.	9905-01355124
PR	21-MAY-1999.	9905-01353533
PR	24-MAY-1999.	9905-01356529
PR	27-MAY-1999.	9905-01360211
PR	27-MAY-1999.	9905-01363032
PR	28-MAY-1999.	9905-01367682
PR	01-JUN-1999.	9905-01372722
PR	04-JUN-1999.	9905-01375528
PR	07-JUN-1999.	9905-01377524
PR	08-JUN-1999.	9905-01380694
PR	10-JUN-1999.	9905-01385400
PR	10-JUN-1999.	9905-01388477
PR	14-JUN-1999.	9905-01391119
PR	16-JUN-1999.	9905-01394452
PR	16-JUN-1999.	9905-01394453
PR	17-JUN-1999.	9905-01394932
PR	18-JUN-1999.	9905-01394954
PR	18-JUN-1999.	9905-01394945
PR	18-JUN-1999.	9905-01394652
PR	18-JUN-1999.	9905-01394611
PR	18-JUN-1999.	9905-01394563
PR	18-JUN-1999.	9905-01397510
PR	18-JUN-1999.	9905-01397503
PR	21-JUN-1999.	9905-01398117
PR	22-JUN-1999.	9905-01398619
PR	23-JUN-1999.	9905-01400353
PR	23-JUN-1999.	9905-01400354
PR	24-JUN-1999.	9905-01400695
PR	28-JUN-1999.	9905-01400823
PR	29-JUN-1999.	9905-01400921





GenCore version 5.1.4-B5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 19:35:14 : Search time 1497 Seconds  
(without alignments)  
3786.520 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 350

Sequence: 1 atgagctcagctacgttcga.....ttgagagcttcacataaaa 350

Scoring table: IDENTITY\_NUCDX

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:  
1: em\_estba:\*  
2: em\_estlum:\*  
3: em\_estlin:\*  
4: em\_estnu:\*  
5: em\_estcov:\*  
6: em\_estcpl:\*  
7: em\_estcro:\*  
8: em\_estc1:\*  
9: qb\_est1:\*  
10: qb\_est12:\*  
11: qb\_estc1:\*  
12: qb\_estc3:\*  
13: qb\_estc4:\*  
14: qb\_estc5:\*  
15: em\_estlum:\*  
16: em\_estlum:\*  
17: qb\_estc1:\*  
18: em\_estlum:\*  
19: em\_estlum:\*  
20: em\_estlum:\*  
21: em\_estlum:\*  
22: em\_estlum:\*  
23: em\_estlum:\*  
24: em\_estlum:\*  
25: em\_estlum:\*  
26: em\_estlum:\*  
27: em\_estlum:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	250.8	71.7	558	12	BF118340 uy93h12.x
C 2	249.2	71.2	442	12	BE548361 UI-M-BH3-
C 3	249.2	71.2	787	12	BC069460 H3075H05-
C 4	249.2	71.2	1040	11	AK012265 Mus muscu
C 5	248.2	70.9	1040	11	AK012586 Mus muscu
C 6	244.4	69.8	420	5	AA960653 ub60c05.s

C 7	241.6	69.0	423	9	AU020489	AU020489
C 8	238.2	68.1	460	9	AU019051	AU019051
C 9	237.2	67.8	607	12	BE549377	BE549377
C 10	225.2	64.3	383	5	AA988210	UI-M-BH3-
C 11	225.2	64.3	410	9	AU012872	UI-M-BH3-
C 12	225.2	64.3	410	13	H1288680	UI-M-BH3-
C 13	225.2	64.3	423	13	H1293426	UI-M-BH3-
C 14	225.2	64.3	559	9	A1103164	UI-M-BH3-
C 15	225.2	64.3	584	5	A1231308	UI-M-BH3-
C 16	224.4	64.1	392	5	A1712611	UI-M-BH3-
C 17	223.6	64.9	462	10	AA434801	UI-M-BH3-
C 18	223.6	64.9	472	12	BF395018	UI-M-BH3-
C 19	215.6	61.6	612	9	AA799579	UI-M-BH3-
C 20	207.6	59.3	355	5	A1704461	UI-M-BH3-
C 21	201.2	57.5	412	5	A1339365	UI-M-BH3-
C 22	201.2	57.5	413	10	AA468894	UI-M-BH3-
C 23	201.2	57.5	425	12	BC163526	UI-M-BH3-
C 24	201.2	57.5	462	12	BF939618	UI-M-BH3-
C 25	201.2	57.5	489	5	A1347440	UI-M-BH3-
C 26	201.2	57.5	529	10	AA516903	UI-M-BH3-
C 27	201.2	57.5	551	5	A1758331	UI-M-BH3-
C 28	201.2	57.5	553	9	A1458831	UI-M-BH3-
C 29	201.2	57.5	560	14	BM682266	UI-M-BH3-
C 30	201.2	57.5	575	13	BM653438	UI-M-BH3-
C 31	201.2	57.5	645	9	A1472315	UI-M-BH3-
C 32	201.2	57.5	656	9	A1921608	UI-M-BH3-
C 33	201.2	57.5	804	12	BE744630	UI-M-BH3-
C 34	199.6	57.0	380	10	AW078977	UI-M-BH3-
C 35	199.6	57.0	593	14	B0021517	UI-M-BH3-
C 36	199.6	57.0	623	9	A1719649	UI-M-BH3-
C 37	197.4	55.4	348	9	A1713665	UI-M-BH3-
C 38	197.2	56.3	460	13	H1402360	UI-M-BH3-
C 39	196.4	56.1	303	9	AA667227	UI-M-BH3-
C 40	196.4	56.1	723	9	A1749254	UI-M-BH3-
C 41	196.4	56.1	726	10	AW001185	UI-M-BH3-
C 42	193.2	55.2	429	9	AA045722	UI-M-BH3-
C 43	193.2	55.2	753	13	H1831537	UI-M-BH3-
C 44	193.2	55.2	957	14	B0931566	UI-M-BH3-
C 45	191.2	54.6	294	9	AA866678	UI-M-BH3-

#### ALIGNMENTS

RESULT 1  
BF118340/c  
LOCUS  
DEFINITION  
uy93h12.x1 NC1\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3667175 3'  
similar to YR:092517 092517 RRP-MS/TYPF 3. // mRNA sequence.

ACCESSION  
BF118340  
VERSION  
BF118340.1 GI:10587816

KEYWORDS  
EST.

SOURCE  
house mouse.

ORGANISM  
Mus musculus

REFERENCE  
Mammalia: Eutheria: Rodentia: Sclurognathii: Muridae: Murinae: Mus.

AUTHORS  
1 (bases 1 to 558)

TITLE  
NC1\_CGAP http://www.ncbi.nlm.nih.gov/ncicgap

COMMENT  
Unpublished (1997)

CONTACT  
Other ESTs: uy93h12.y1

REMARKS  
Contact: Robert Strausberg, Ph.D.

REMARKS  
Tissue Procurement: Jolnar Hennighausen Ph.D., Robin Humphreys

REMARKS  
CDNA Library Preparation: Life Technologies, Inc.

REMARKS  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

REMARKS  
Clone distribution: NC1-CGAP clone distribution information can be

REMARKS  
found through the I.M.A.G.E. Consortium/LLNL at:

REMARKS  
image.llnl.gov/image/llnl/resources.shtml

REMARKS  
MG1:1427943







## CDS

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAAGACCACTCAGACCCTCTTTTGTATTTTGTNN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAAGATTTCGAGTGATTTAATAATATCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

```
/gene="Rbpms"
```

**TITLE** Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
**JOURNAL** Genome Res. 10 (10), 1617-1630 (2000)  
**MEDLINE** 20499374



REFERENCE 1 (bases 1 to 420)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Stepien, M., Tan, F., Underwood, R., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:904588  
 Seg primer: -40m3 fwd. ET from Amersham  
 High quality sequence stop: 159.

FEATURES  
 source  
 1. 420  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1382120"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="Female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 116 a 71 c 132 g 101 t

ORIGIN

Query Match 69.8%; Score 244.4; DB:9; Length 420;  
 Best Local Similarity 78.3%; Pred. No. 6.8e-28;  
 Matches 274; Conservative 30; Mismatches 41; Indels 5; Gaps 3;

QY 1 ATAGGCTCAAGTACCTGACCTTACCCAGTACCTGAGTGTGGCCCCGTTACCTC 60  
 DB 359 ATAGGCTCAAGTACCTGACCTTACCCAGTACCTGAGTGTGGCCCCGTTACCTC 300  
 QY 61 TGTACCCAGGAGTGTAGCCCTGCTCTCTCTGCGGCTTACCTACCCGCTT 120  
 DB 299 TGTACCCAGGAGTGTAGCCCTGCTCTCTCTGCGGCTTACCTACCCGCTT 240  
 QY 121 CACTGATGCCAGAGGAGGCTGTATACATGCTAGGNNNNNACNCTGACNNNTGAGAC 180  
 DB 239 CACTGATGCCAGAGGAGGCTGTATACATGCTAGGNNNNNACNCTGACNNNTGAGAC 180  
 QY 181 TNCCTNACGACATGATGATGCTGCTGGAATGAGCTGACAGATCAACTACTG 240  
 DB 179 TTTCTTACGCA---GATTAGATTTGTCTCTGGAATGATGTTGTCATATCAACTACTG 240  
 QY 241 CACAGCANCNAGAGGAGGCTGNNNGACAGATGATATANNACNNACTACTATNCTT 300  
 DB 122 CACAGCAGGAGGAGGAGGAGGCTGCGTG-TTGAAATTTATTTAGAACCCCTT-CCTGCCCCCTT 65  
 QY 301 TTGCAAAAGNAACTTTCTTTGANAATNTTTAGAGATTTCAATAAAAA 350  
 DB 64 TTGCAAAAGNAACTTTCTTTGANAATNTTTAGAGATTTCAATAAAAA 15

RESULT 7  
 A0020489/c 423 bp mRNA linear EST 19-OCT-1998

DEFINITION A0020489 Mouse eight-cell stage embryo cDNA Mus musculus cDNA clone  
 J0533602 3', mRNA sequence.  
 ACCESSION A0020489  
 VERSION A0020489.1 GI:3376073  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 423)  
 AUTHORS Liang, Y., Kargul, G.J., Sharara, R. and Doi, H.  
 TITLE Systematic analyses of genes expressed in eight-cell stage mouse embryos (The ERATO/Doi project at Wayne State University) (Ko, M.S.H. et al.)  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Hirotumi Doi  
 Doi Biosymmetry Project, ERATO  
 Japan Science and Technology Corporation (JST)  
 MBI Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan  
 Email: hdoe@ia.jst.go.jp.

FEATURES  
 source  
 1. 423  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="J0533602"  
 /clone\_lib="Mouse eight-cell stage embryo cDNA"  
 /dev\_stage="eight-cell stage embryo"  
 /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5].  
 TGTACCAATCTGAAGTGGAGAGGCGGCGGAGATGTTTTTTTTTTTTTTTTTTT  
 T 3'; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
 constructed and normalized by Bento Soares and M. Fatima  
 Bonaldo."

BASE COUNT 119 a 73 c 132 g 98 t 1 others

ORIGIN

Query Match 69.0%; Score 241.6; DB:9; Length 423;  
 Best Local Similarity 78.5%; Pred. No. 1.9e-27;  
 Matches 270; Conservative 30; Mismatches 39; Indels 5; Gaps 3;

QY 7 TCAAGTACCTGACCTTACCCAGTACCTGAGTGTGGCCCCGTTACCTCTGAC 66  
 DB 352 TCAAGTACCTGACCTTACCCAGTACCTGAGTGTGGCCCCGTTACCTCTGAC 293  
 QY 67 CAGCGAGTGTAGCGCTGCTCTCTCTGCGGCTTACCTACCCGCTTACTGC 126  
 DB 292 CAGCGAGTGTAGCGCTGCTCTCTCTGCGGCTTACCTACCCGCTTACTGC 223  
 QY 127 ATGCCAGAGGAGGAGGCTGTATACATGCTAGGNNNNNACNCTGACNNNTGAGAC 186  
 DB 232 ATGCCAGAGGAGGAGGCTGTATACATGCTAGGNNNNNACNCTGACNNNTGAGAC 186  
 QY 187 AGCAGATGATGATGATGCTGCGAATGAGCTGACAGATCAACTACTGACAGAC 246  
 DB 172 AGCA---GATTAGATTTGTCTGGAATGATGTTGTCATATCAACTACTGACAGAC 116  
 QY 247 ANCAAGCAGGCTTNNCAGACAGATGATATANNACNAACTACTATNCTTTGCAA 306  
 DB 115 AGCAGAGGAGGAGGAGGCTGCGTG-TTGAAATTTATTTAGAACCCCTT-CCTGCCCCCTTTGCAA 58  
 QY 307 ANCAAACTTTCTTTGANAATNTTTAGAGATTTCAATAAAAA 350  
 DB 57 AAAAAAACTTTCTTTGANAATNTTTAGAGATTTCAATAAAAA 14

RESULT 8

LOCUS	AAU019051/c		460 bp	mRNA	linear	EST 19-Oct-1998
DEFINITION	AAU019051 Mouse eight-cell stage embryo cDNA Mus musculus cDNA clone J0512C09.3, mRNA sequence.					
ACCESSION	AAU019051					
VERSION	AAU019051.1		GI:3374635			
KEYWORDS	EST.					
SOURCE	Mus musculus house mouse.					
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	1 (bases 1 to 460) Ko,M.S.H., Kitchen,J.R., Wang,X., Threatt,T.A., Sim,T.T., Depaulis,G.E., Liang,Y.Y., Karguly,G.J., Sharrara,R. and Del.H.					
AUTHORS	Systematic analyses of genes expressed in eight-cell stage mouse embryos (The ERATO/Dol Project at Wayne State University) (Ko ,M.S.H., et al.)					
JOURNAL	Unpublished (1998)					
COMMENT	Contact: Hirofumi Dol Dol Bioasymmetry Project, ERATO Japan Science and Technology Corporation (JST) Mitsui Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan Email: hded@lon.jst.go.jp. Location/Organism: 1. .460 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="j0512C09" /dew-lib="mouse eight-cell stage embryo cDNA" /dex_stage="eight-cell stage embryo" /note="Organ: mammary gland; Vector: pT7T3D-Pac (pharmacia ) with a modified polylinker; site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5]. TCTTACCATCTGCACTGGAGCGGCCGCACAAGCTTTTTTTTTTTTTTTTTTTTTT T 3' I : double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Felima Ronaldo."					
BASE COUNT	120 a 80 c 143 g 116 t 1 others					
ORIGIN						
Query Match	68.1% Score 238.2; PB 5; Length 460;					
Best Local Similarity	78.9% Pred. No. 7.4e-27					
Matches 277; Conservative	30; Mismatches 38; Indels 6; Gaps 4;					
Dh	1 ATGACGTGACAGTAGTCGACTTTAACCCAGTAGCCCTTG-AAGTGTGGCGCCGCTAGCCCT 59					
Dh	359 ATGAGCTCACAGTACCTCCACTTAACCCAGTAGCCCTGAAGAATTGGGCCCGCTACCCCT 300					
Dh	60 CTGTACCAGCAGGAGTATGAGCCCTGCTCTTCCTCTCTCCGCGGCTTTTACGACACCCGCT 119					
Dh	239 CTGTACCCAGCAGGAGTATGAGCCCTGCTCTTCCTCTCTCCGCGGCTTTTACGACACCCGCT 240					
Dh	120 TCACTGCATGGCCAGAGGAGCNCGTATACATGCTCTAGGNNNNACNCGCTGACNNNTGGANA 179					
Dh	239 TCACGTGCATGCCCAAGTCTTCTCTCCAGAGCAAGCAAGCCCAACACACGTGCTTTTGTGCA 180					
Dh	180 CTNCTNCGACCATGNATGATNGCGTGCATGCGGGAATGAGCCGTCNCGACGATCAACACTACT 239					
Dh	179 GTTCTTCAGCA---GATTAGATTGTTCTCTGCGAATGTGCTTTGTCAATATCAACTACT 123					
Dh	240 GCAGACCANNAAGGAGCACTTNNGACAGACATGATATTANNAAAGCNACTACTTATTTNCT 299					
Dh	122 GCAGACCAAGCAGAGAGGAGCTCCCGTG-TTGAATTTATTTAGCAACCCCT-CGCTCCGCT 65					
Dh	300 TTTCGCAANGANAAGCTNTTCTTTGGANAATNTTTGAGAGATTTTCAATATAAA 350					
Dh	64 TTTCGCAAAAAAAGCTAATTTCTGATATATTGTGAGAGCTTTTCATATAAAA 14					

RESULT 9	BE945377/c	607 bp	mrna	linear	EST 03-oct-2000
LOCUS	BE945377				
DEFINITION	U1-M-BH3-avl-c-01-0-U1.s1 NIH-BMAP_M_S4		Mus musculus cDNA clone		
ACCESSION	U1-M-BH3-avl-c-01-0-U1.3		mrna sequence.		
VERSION	BE945377				
KEYWORDS	BE945377.1	GI:10527136			
SOURCE	EST.				
ORGANISM	House mouse.				
REFERENCE	Mus musculus				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
TITLE	1 (bases 1 to 607)				
	Ronald M.F., Lennon G. and Soares M.B.				
	Normalization and subtraction: two approaches to facilitate gene discovery				
JOURNAL	Genome Res.	6 (9)	791-806	(1996)	
MEDLINE	9704447				
COMMENT	Contract: China, H				
	National Institute of Mental Health				
	6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD				
	20892-9643, USA				
	Tel: 301 443 1705				
	Fax: 301 443 5850				
	Email: m3estmail.nih.gov				
	The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the brain stems tissue cDNA library preparation: M.B. Soares Lab clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements				
	Seq primer: M13 forward				
	POLYA-Yes.				
FEATURES	Location/Qualifiers				
SOURCE	1..607				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="U1-M-BH3-avl-c-01-0-U1"				
	/clone_lib="NIH-BMAP_M_S4"				
	/dex_stage="27-32 days"				
	/lab_host="DH10B (Life Technologies)"				
	/note="Vector: pTZ193D-pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; The NIH-BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH-BMAP_M_S2, NIH-BMAP_M_S3.3, NIH-BMAP_M_S3.2, NIH-BMAP_M_S3.1, NIH-BMAP_M_S2, NIH-BMAP_M_S1. The subtracted library (NIH-BMAP_M_S4) was constructed as follows: PCR-amplified cDNA inserts from NIH-BMAP_M_S3.3, NIH-BMAP_M_S3.2, and NIH-BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH-BMAP_M_S3.3, NIH-BMAP_M_S3.2, and NIH-BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH-BMAP_M_S4 library. This procedure has been previously described (Ronald M. Soares and Soares, Genome Research				

6:791-806, 1996)  
TAG LIB=NIH BMAP M.S4  
TAG TISSUE=brain-stems  
TAG-SFO=TCATC\*

BASE COUNT 149 a 111 c 177 g 169 t 1 others

Query Match 67.8%; Score 237.2; DB 12; Length 607;

Best Local Similarity 78.9%; Pred. No. 1.5e-26; Mismatches 38; Indels 6; Gaps 4;

Matches 276; Conservative 30; Mismatches 38; Indels 6; Gaps 4;

Db 1 ATGAGCTCAGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
373 ATGAGCTCAGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
61 TGTACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
313 TGTACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
121 CACTGATGCCAGAGGAGCTGTATACATGCTAGGNNNNACNCTGACNNNTGAGAAC  
253 CACTGATGCCAGAGGAGCTGTATACATGCTAGGNNNNACNCTGACNNNTGAGAAC  
181 TGTACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
194 TGTACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
241 CAGACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
137 CAGACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
301 TTGCAAGAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
79 TTGCAAGAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC

RESULT 10  
AA998210/c 383 bp mRNA linear EST 04-JUN-1999

LOCUS AA998210  
DEFINITION UI-R-CO-1b-c-06-0-UI.s1 UI-R-CO Rattus norvegicus cDNA clone  
ACCESSION AA998210  
VERSION AA998210.1 GI:4288286

KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 383)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477

COMMENT On Jun 5, 1998 this sequence version replaced gi:3188861.  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565

Email: msosares@blue.weeg.uiowa.edu  
The sequence tag present in the cDNA between the NotI site and the  
oligo-dT track served to identify it as a clone from the normalized  
adult heart library. cDNA library preparation: M. Fatima Bonaldo,  
Ph.D. Heart distribution: clones will be available through Research  
Genetics This clone is also available through the I.M.A.G.E.  
Consortium at LHM (info@image.llnl.gov). IMAGE ID=1774034  
Seq primer: M13 Forward  
POLYA=No.

FEATURES  
Location/Qualifiers  
1..383

/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-CO-1b-c-06-0-UI"  
/clone\_lib="UI-R-CO"  
/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-CO  
library is a subtracted library derived from the UI-R-A1  
and UI-R-E1 libraries. The UI-R-A1 library consisted of a  
mixture of individually tagged normalized libraries  
constructed from rat placenta, adult lung, brain, liver,  
kidney, heart, spleen, ovary, and muscle. The UI-R-E1  
library consisted of a mixture of individually tagged  
normalized libraries constructed from 8, 12 and 18-day  
embryo. The tag is a string of 3-5 nucleotides present  
between the Not I site and the oligo-dT track which  
allows identification of the library of origin of a clone  
within the mixture. The subtracted library (UI-R-CO) was  
constructed as follows: PCR amplified cDNA inserts from a  
pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had  
been derived was used as a driver in a hybridization with  
the pooled UI-R-A1 and UI-R-E1 library in the form of  
single-stranded circles. The remaining single-stranded  
circles (subtracted library) was purified by  
hydroxyapatite column chromatography, converted to  
double-stranded circles and electroporated into DH10B  
bacteria (Life Technologies) to generate the UI-R-CO  
library. This procedure has been previously described  
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,  
1996)."

BASE COUNT 103 a 68 c 111 g 101 t

Query Match 64.3%; Score 225.2; DB 9; Length 383;  
Best Local Similarity 75.4%; Pred. No. 7e-25;  
Matches 264; Conservative 29; Mismatches 48; Indels 9; Gaps 3;

Db 1 ATGAGCTCAGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
371 ATGAGCTCAGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
61 TGTACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
311 TGTACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
121 CACTGATGCCAGAGGAGCTGTATACATGCTAGGNNNNACNCTGACNNNTGAGAAC  
251 CACTGATGCCAGAGGAGCTGTATACATGCTAGGNNNNACNCTGACNNNTGAGAAC  
181 TGTACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
191 TGTACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
241 CAGACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
134 CAGACCCAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
301 TTGCAAGAGGAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGGTACCCCTC  
75 T-----GCNAAACTATTTCTTGATGATATTGAGAGGATTCATATAAAA 31

RESULT 11  
A1712872/c 410 bp mRNA linear EST 08-JUN-1999

LOCUS A1712872  
DEFINITION UI-R-A1-aaf-e-04-0-UI.s2 UI-R-A1 Rattus norvegicus cDNA clone  
ACCESSION A1712872  
VERSION A1712872.1 GI:5016672

KEYWORDS EST.  
SOURCE Norway rat.









```
|||||
Db 252 CACGCGATGCCAGTGTCTCTCCAGAGCAAGGCCACACACCGCTGTCTTTCCAC 193
      181 TNCNCAGCATGNATGATNGGTCGTCGGATGAGCTGCTGACGATCATCTAC 240
      192 TTTCTTCAGCAAT--TACATTTGTCTCTGGATGTGTTGTGACGTATCAACCCAGCTG 136
QY 241 CAGACGACNAGGAGGACCTTNGAGACAGATGTATPANNACGACTTATTCCTT 300
      135 CAGACGACGAGAGGAGAC-TCCCATGTTGAATTTATTAGACAGCTCCCTGCCCTTT 77
Db 301 TTGCAANAGNAACTNTTCTTGANAATNTTTGAGAGATTTCATAAAAA 350
      76 T-----GCAAAACTATTTCTGTATGATATTGAGAGATTTCATAAAAA 32
```

## RESULT 14

```
LOCUS A1103164 559 bp mRNA linear EST 31-JAN-1999
DEFINITION EST2.2453 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
ACCESSION A1103164
VERSION A1103164.1 GI:3707736
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 559)
AUTHORS Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@fcr.org
Seq primer: M13-21.
```

## FEATURES

## source

```
1..559
/organism="Rattus sp."
/db_xref="ATCC (inhost):2024504"
/db_xref="taxon:10118"
/clone="KEMBX96"
/clone_1lb="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
Seq primer: M13-21.
```

## BASE COUNT

144 a 108 c 156 g 141 t

## ORIGIN

Query Match 64.3%; Score 225.2; DB 9; Length 559;  
Best Local Similarity 75.7%; Pred. No. 1.1e-24;  
Matches 265; Conservative 28; Mismatches 48; Indels 9; Gaps 3;

```
QY 1 ATGAGCTCAGAGTACCTGCACTTACGCCAGTACGCTGAGAGTGTGGGCCCGGACCTTC 60
      354 ATGAGCTCAGAGTACCTGCACTTACGCCAGTACGCTGAGAGTGTGGGCCCGGACCTTC 295
Db 61 TGTACCCAGGAGATTTAGCGCTGCTCTCTCTGCGGCTTTGACGTACCGCGCTT 120
      294 TGTACCCAGGAGATTTAGCGCTGCTCTCTCTGCGGCTTTGACGTACCGCGCTT 235
QY 121 CACTGCATGCCAGAGGCGCTGATACATGCTAGCGNNNNACGCTGACNNNTGANAAC 180
      234 CACTGCATGCCAGAGGCGCTGCTCTCTGCGAGGCAAGGCCACACACGCTGTCTTTGTCAC 175
Db 181 TNCNCAGCATGNATGATNGGTCGTCGGATGAGCTGACGATCATCTAC 240
      174 TTTCTTCAGCAAT--TACATTTGTCTCTGGATGTGTTGTGACGTATCAACCCAGCTG 118
```

```
QY 241 CAGACGACNAGGAGGACCTTNGAGACAGATGTATPANNACGACTTATTCCTT 300
      117 CAGACGACGAGAGGAGGACCTCCCATGTTGAATTTATTAGCAAGCT--TCCCTGCCCTT 60
QY 301 TTGCAANAGNAACTNTTCTTGANAATNTTTGAGAGATTTCATAAAAA 350
      59 TTGCA--AAAACTATTTCTGTATGATATTGAGAGATTTCATAAAAA 14
```

## RESULT 15

```
LOCUS A1231308 584 bp mRNA linear EST 31-JAN-1999
DEFINITION EST227956 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
ACCESSION A1231308
VERSION A1231308.1 GI:3815188
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 584)
AUTHORS Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@fcr.org
Seq primer: M13-21.
```

## FEATURES

## source

```
1..584
/organism="Rattus sp."
/db_xref="ATCC (inhost):2038893"
/db_xref="taxon:10118"
/clone="KEMD142"
/clone_1lb="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
Seq primer: M13-21.
```

## BASE COUNT

147 a 114 c 174 g 145 t

## ORIGIN

Query Match 64.3%; Score 225.2; DB 9; Length 584;  
Best Local Similarity 75.4%; Pred. No. 1.2e-24;  
Matches 264; Conservative 29; Mismatches 48; Indels 9; Gaps 3;

```
QY 1 ATGAGCTCAGAGTACCTGCACTTACGCCAGTACGCTGAGAGTGTGGGCCCGGACCTTC 60
      354 ATGAGCTCAGAGTACCTGCACTTACGCCAGTACGCTGAGAGTGTGGGCCCGGACCTTC 295
Db 61 TGTACCCAGGAGATTTAGCGCTGCTCTCTCTGCGGCTTTGACGTACCGCGCTT 120
      294 TGTACCCAGGAGATTTAGCGCTGCTCTCTCTGCGGCTTTGACGTACCGCGCTT 235
QY 121 CACTGCATGCCAGAGGCGCTGATACATGCTAGCGNNNNACGCTGACNNNTGANAAC 180
      234 CACTGCATGCCAGAGGCGCTGCTCTCTGCGAGGCAAGGCCACACACGCTGTCTTTGTCAC 175
Db 181 TNCNCAGCATGNATGATNGGTCGTCGGATGAGCTGACGATCATCTAC 240
      174 TTTCTTCAGCAAT--TACATTTGTCTCTGGATGTGTTGTGACGTATCAACCCAGCTG 118
QY 241 CAGACGACNAGGAGGACCTTNGAGACAGATGTATPANNACGACTTATTCCTT 300
      117 CAGACGACGAGAGGAGAC-TCCCATGTTGAATTTATTAGCAAGCTCCCTGCCCTTT 59
Db 301 TTGCAANAGNAACTNTTCTTGANAATNTTTGAGAGATTTCATAAAAA 350
```

Mon May 12 08:00:25 2003

Db 58 T-----GCAAAACCTATTCTTCTGATGATATTGAGAGATTTCATTAATAA 14

Search completed: May 9, 2003, 20:39:43  
Job time : 1504 secs

us-09-880-711-328.rst

GenCore version 5.1.4.p3.4578  
Copyright (c) 1993 - 2003 Computer Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 9, 2003, 19:38:48 ; Search time 67 Seconds  
(without alignments)  
1802.043 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 350  
Sequence: 1 atgagcgcacgtacctgca.....ctgaagatcttcaataaaa 150

Scoring table: IDENTITY\_NUCDX  
Gapop 10.0 , Gapext 1.0

Searched: 141362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA: \*  
1: /cgn2\_6/pdata1/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/pdata1/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/pdata1/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/pdata1/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/pdata1/1/ina/PC105\_COMB.seq: \*  
6: /cgn2\_6/pdata1/1/ina/Backfiles...seq: \*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.8	17.4	7218	1	US-08-232-463-14 Sequence 14, Appl
2	56.4	16.1	9060	4	US-08-378-313-20 Sequence 20, Appl
3	54.8	15.7	2035	4	US-08-960-780-10 Sequence 10, Appl
4	54.8	15.7	2035	4	US-09-073-898-10 Sequence 10, Appl
5	54.2	15.5	751	4	US-09-221-017B-18 Sequence 18, Appl
6	54.2	15.5	3619	4	US-08-377-503-11 Sequence 11, Appl
7	54.2	15.5	3619	4	US-08-178-019-11 Sequence 11, Appl
8	53.4	15.3	2200	1	US-08-592-125-102 Sequence 102, App
9	53.4	15.3	7452	3	US-08-592-500-11 Sequence 11, Appl
10	53.4	15.3	7452	3	US-08-195-005-11 Sequence 11, Appl
11	53.4	15.3	7452	5	PCT-US94-07644A-1 Sequence 1, Appl
12	52.4	15.0	1297	1	US-08-370-975B-12 Sequence 12, Appl
13	52.4	15.0	2593	4	US-08-936-165A-153 Sequence 153, App
14	52.4	15.0	5150	4	US-09-068-140A-14 Sequence 14, Appl
15	52.4	15.0	26764	1	US-08-370-975B-11 Sequence 11, Appl
16	52.2	14.9	1425	1	US-08-353-550-77 Sequence 7, Appl
17	52.2	14.9	1425	2	US-08-551-687-77 Sequence 7, Appl
18	52.2	14.9	1512	2	US-07-938-154-110 Sequence 10, Appl
19	52.2	14.9	1512	5	PCT-US91-02311-10 Sequence 10, Appl
20	52.2	14.9	1610	4	US-09-059-769-19 Sequence 19, Appl
21	52.2	14.9	1819	4	US-09-453-702B-188 Sequence 188, App
22	52.2	14.9	3031	4	US-08-664-962B-33 Sequence 3, Appl
23	52.2	14.9	3031	4	US-09-311-743-33 Sequence 3, Appl
24	51.6	14.7	3797	4	US-08-936-165A-107 Sequence 107, App
25	51.4	14.7	4853	4	US-08-881-450A-22 Sequence 22, Appl
26	51.2	14.6	3872	5	PCT-US93-12169-11 Sequence 11, Appl
27	51	14.6	3116	1	US-08-149-103-22 Sequence 2, Appl

C	28	51	14.6	3116	1	US-08-451-883-2	Sequence 2, Appl
C	29	50.8	14.5	859	4	US-09-247-373B-47	Sequence 47, Appl
C	30	50.8	14.5	977	1	US-08-017-522A-1	Sequence 1, Appl
	31	50.8	14.5	977	6	5215895-2	Patent No. 5215895
	32	50.8	14.5	1100	1	US-07-949-516A-1	Sequence 1, Appl
	33	50.8	14.5	1100	2	US-08-814-459-1	Sequence 1, Appl
	34	50.8	14.5	1100	3	US-09-122-525-1	Sequence 1, Appl
	35	50.8	14.5	2073	4	US-09-134-001C-1731	Sequence 1731, Ap
	36	50.8	14.5	2167	3	US-08-368-704C-52	Sequence 52, Appl
	37	50.8	14.5	2174	3	US-08-258-287B-54	Sequence 54, Appl
	38	50.6	14.5	2119	4	US-09-240-519-7	Sequence 7, Appl
C	39	50.6	14.5	14104	4	US-08-961-527-34	Sequence 34, Appl
	40	50.4	14.4	2175	2	US-08-389-564B-4	Sequence 4, Appl
	41	50.4	14.4	2175	3	US-08-466-047B-4	Sequence 4, Appl
	42	50.4	14.4	2437	1	US-07-795-859B-5	Sequence 5, Appl
	43	50.4	14.4	2437	1	US-08-457-616-5	Sequence 5, Appl
	44	50.4	14.4	2437	4	US-09-235-518-1	Sequence 1, Appl
C	45	50.4	14.4	8931	3	US-09-028-934-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: LAWRENCE, F.  
APPLICANT: SCHIEFLINGER, F.  
APPLICANT: FAULKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESS: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0295  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300-6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/PROCKET NUMBER: 30472/114 1MMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZapTc-r1s  
US-08-232-463-14  
Query Match: 17.4%; Score 60.8; DB 1; Length 7218;

Best Local Similarity 5.1%; Pred. No. 5.5;  
Matches 17; Conservative 185; Mismatches 132; Indels 0; Gaps 0;

QY 6 CTCACAGTACTGACCTTACCCAGCCAGTACGAGTGTGGCCCGCTACCTGTAC 65  
Db 1140 YY 1199  
QY 66 CCAGCGAGTACGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 125  
Db 1200 YY 1259  
QY 126 CATGCCAGAGGCTGATACATGCTTACGAGGCTTACCTACCTACCTACCT 185  
Db 1260 YY 1319  
QY 186 CAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1245  
Db 1320 YY 1379  
QY 246 CANCAAGGAGCTTNNAGACAGATGATGATGATGATGATGATGATGATGAT 1305  
Db 1380 YY 1439  
QY 306 AANGAAGTCTTCTTGANATNTTGGAGAT 339  
Db 1440 AATCTCTCTATCTCTTAACTACTGATGAT 1473

## RESULT 2

US-08-378-313-20/c  
Sequence 20, Application US/08378313  
Patent No. 620781  
GENERAL INFORMATION:  
APPLICANT: THEOLOGIS, ATHANASIOS  
TITLE OF INVENTION: CONTROL OF FRUIT RIPENING THROUGH  
TITLE OF INVENTION: GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOENSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/378, 313  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/862,493  
FILING DATE: 02-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 29190-20002.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 856-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9060 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS

LOCATION: join(2704..2880, 2968..3099, 3183..3344, 3810  
LOCATION: ..4376, 4463..4903)  
US-08-378-313-20

Query Match 16.1%; Score 56.4; DB 4; Length 9060;  
Best Local Similarity 40.7%; Pred. No. 51;  
Matches 79; Conservative 29; Mismatches 86; Indels 0; Gaps 0;

QY 157 NNNNACNCTGACNNNTGACNNCTNCAGACATGATGATGATGATGATGAT 216  
Db 7502 ATTTCCAGATATTCATTTATTTATTTATTTATTTATTTATTTATTTATTT 7443  
QY 217 NGCCTGACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 276  
Db 7442 AGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7383  
QY 277 TANNACNNACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 336  
Db 7382 TGTCACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 7323  
QY 337 GATTTCATTAATAA 350  
Db 7322 AATTCAAAATGAAAA 7309

## RESULT 3

US-08-960-780-10  
Sequence 10, Application US/08960780  
Patent No. 620435  
GENERAL INFORMATION:  
APPLICANT: Feltelson, Jerald S.  
APPLICANT: Schepf, H. Ernest  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Stockhoff, Brian A.  
APPLICANT: Schmets, James  
APPLICANT: Loewer, David  
APPLICANT: Dullum, Charles Joseph  
APPLICANT: Muller-Cohn, Judy  
APPLICANT: Stamp, Lisa  
TITLE OF INVENTION: No. 6204435e1 Pesticidal Toxins and Nucleotide  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: US  
ZIP: 32606-6669  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/960,780  
FILING DATE: 30-OCT-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/029,848  
FILING DATE: 30-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: MA-708  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 352-375-8100  
TELEFAX: 352-372-5800  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2035 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single



REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 751 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORPHYROMOMAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...751  
US-09-221-01/B-18

Query Match 15.5%; Score 54.2; DB 4; Length 751;  
Best Local Similarity 42.7%; Pred. No. 5.3;  
Matches 85; Conservative 30; Mismatches 83; Indels 1; Gaps 1;

QY 136 GGCNCTGTATACATGCTAGGNNNNACNCCGTGACNNNTGACNCTGACACATGN 195  
DB 404 GGAGCTTCGTTCACGCTAGCGCTGTTCTTGACATTCGCAACATCATTCTGCG 345  
QY 196 ATGATNGCGTCTGCGAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 255  
DB 344 TTATCTCCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 286  
QY 256 AGCTTNGACAGAAATGATTTANNAACNACTACTTATNCTTTTGCANAAGAACT 315  
DB 285 AACGGGATTAAGAAGATTCGATTCACACAGAGATGGGGTGGGAGTTGGGTTGCA 226  
QY 316 NTCTTGAATTTTGA 334  
DB 225 AACGTGTGACTGATGA 207

RESULT 6  
US-08-377-503-1/c  
Sequence 1, Application US/08377503  
Patent No. 6322999

GENERAL INFORMATION:  
APPLICANT: Kamboj, Rajender  
APPLICANT: Elliot, Candace E.  
APPLICANT: Nutt, Stephen L.  
TITLE OF INVENTION: Kainate-Binding, Human CNS Receptors of  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: Suite 500, 3000 K Street  
CITY: N.W.  
STATE: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/377,503  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/945,210  
FILING DATE: 17-SEP-1992  
ATTORNEY/AGENT INFORMATION:

NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16777/192  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3619 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 16..108  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 109..2772  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 16..2772  
US-08-377-503-1

Query Match 15.5%; Score 54.2; DB 4; Length 3619;  
Best Local Similarity 40.1%; Pred. No. 41;  
Matches 75; Conservative 29; Mismatches 83; Indels 0; Gaps 0;

QY 148 ATGCTAGGNNNNACNCCGTGACNNNTGACNCTGACACATGATGATGCGGCGG 207  
DB 1763 ATGAGAGAGAGACACACCTGACCCCGAGTAGGAGAGAGACATCATCATGCT 1704  
QY 208 CTGGCAATGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 267  
DB 1703 GGGGACAGGGGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1644  
QY 268 AGAATGTATTTANNAACNACTACTTATNCTTTTGCANAAGAACTNTCTTGANA 327  
DB 1643 AGATGCTACACACCAAGTGTATGAGGCTTGGAGAGAGTGTAGTGTCTTCGACA 1584  
QY 328 TTTTGA 334  
DB 1583 TCGGTGA 1577

RESULT 7  
US-08-178-019-1/c  
Sequence 1, Application US/08178019  
Patent No. 6333161

GENERAL INFORMATION:  
APPLICANT: Kamboj, Rajender  
APPLICANT: Elliot, Candace E.  
APPLICANT: Nutt, Stephen L.  
TITLE OF INVENTION: Kainate-Binding, Human CNS Receptors of  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: Suite 500, 3000 K Street  
CITY: N.W.  
STATE: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/178,019  
FILING DATE: 06-JAN-1994  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Beall, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 15777/227  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5359  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3619 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 16..108  
FEATURE:  
NAME/KEY: mal-peptide  
LOCATION: 109..2772  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 16..2772  
US-08-178-019-1

Query Match 15.5%; Score 54.2; DB 4; Length 3619;  
Best Local Similarity 40.1%; Pred. No. 41;  
Matches 75; Conservative 29; Mismatches 83; Indels 0; Gaps 0;

QY 148 ATGCTAGGNNNNCCAGCNCCTGACNNNTGAGNACTGTCAGCATGATGATNAGTCN 207  
DB 1763 ATGACGACAGACGACAGAGTGACCCGAGGTAGCGCAGAGACATACATCGCATGTCT 1704  
QY 208 CTGGAGATGCGCTGCGACAGATCAACTCTGCGACGCCANCGAGGAGCTTTNCGACG 267  
DB 1703 GGGGACAGCGGATGAGGAGGAGGAGGAGCGTGGGCTTGCTGCATTTGGCTTCGATAC 1644  
QY 268 AGATGCTATTANNACGNCNCTGATTCCTTTTGCAANACNACTGTTCTTGANAA 327  
DB 1643 AGGATGCTACACGCAAGTGTCTGAAGGCTTGGACAGCTGATGCTTCTCTCGACACA 1584  
QY 328 TTTTGA 334  
DB 1583 TCGGTGA 1577

RESULT 8  
US-08-592-126-102  
Sequence 102, Application US/08592126  
Patent No. 5821091  
GENERAL INFORMATION:  
APPLICANT: Gregory Doljancov  
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
NUMBER OF SEQUENCES: 151  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dellinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,126  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles R.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 102:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2200 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: G89con.seq  
US-08-592-126-102

Query Match 15.3%; Score 53.4; DB 1; Length 2200;  
Best Local Similarity 39.1%; Pred. No. 30;  
Matches 84; Conservative 30; Mismatches 101; Indels 0; Gaps 0;

QY 121 GACTGATAGGAGAGGCMCTATACATGCTAGGNNNNCCAGCNCCTGACNNNTGAGAC 180  
DB 1817 CCTGCTGAGGGGCAAGGCGCGCTCTGCGCCAGAGCGCAACTGAGAGAGCATCTC 1876  
QY 181 TWTNAGACATGATGATGTCGTCGCGATGCGGATGCGCTGACAGATGCAACTACTG 240  
DB 1877 GCTTCCTTACAGCAATGAGAGCGCATGAGAGCGCATGAGAGCGCTTCATGAGCCCTCTC 1936  
QY 241 CGAGCAATCGAGGAGCTTNNAGACGATGATATTANNACGNCNCTGATTCCTT 300  
DB 1937 CCAGCCGAGCGGCGCCCGCAGGACTGTCTGCACTTACGCGGCCCTTCCCTCTCTCG 1996  
QY 301 TTGCAANACNACTTTCTTGANAAATTTTGA 335  
DB 1997 TTACATAGATACACATGCGTCAAGCAGCGCTGAG 2031

RESULT 9  
US-08-592-500-1  
Sequence 1, Application US/08592500  
Patent No. 6005085  
GENERAL INFORMATION:  
APPLICANT: Lanza, Francois  
APPLICANT: Phillips, David R.  
TITLE OF INVENTION: Cazenave, Jean-Pierre  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: 375 Bytown Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,500  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/089,455  
FILING DATE: 09-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: How, Karen B.  
REGISTRATION NUMBER: 29,684  
REFERENCE/DOCKET NUMBER: 12418-28

```

1 NAME/KEY: misc_signal
2 LOCATION: 1365..1372
3 OTHER INFORMATION: /function="Ets-1 cis-acting
4 OTHER INFORMATION: sequences"
5 OTHER INFORMATION: /label=Ets-1
6 FEATURE:
7 NAME/KEY: repeat_region
8 LOCATION: 6133..6440
9 OTHER INFORMATION: /rpt_type="other"
10 OTHER INFORMATION: /label=Alu
11 FEATURE:
12 NAME/KEY: misc_signal
13 LOCATION: 5610..5615
14 OTHER INFORMATION: /standard_name="Polyadenylation
15 OTHER INFORMATION: signal sequence"
16 FEATURE:
17 NAME/KEY: misc_signal
18 LOCATION: 6966..6871
19 OTHER INFORMATION: /standard_name="Polyadenylation
20 OTHER INFORMATION: signal sequence"
21 FEATURE:
22 NAME/KEY: misc_signal
23 LOCATION: 7224..7229
24 OTHER INFORMATION: /standard_name="Polyadenylation
25 OTHER INFORMATION: signal sequence"
26 FEATURE:
27 NAME/KEY: misc_signal
28 LOCATION: 7358..7363
29 OTHER INFORMATION: /standard_name="Polyadenylation
30 OTHER INFORMATION: signal sequence"
31 FEATURE:
32 NAME/KEY: misc_feature
33 LOCATION: 1..7452
34 OTHER INFORMATION: /standard_name="Nucleotide
35 OTHER INFORMATION: sequence containing the human cpv gene"
36 US-08-592-500-1
37
38 Query Match 15.3%; Score 53.4; DB 3; Length 7452;
39 Best Local Similarity 39.2%; Fred. No. 1.5e+02;
40 Matches 78; Conservative 30; Mismatches 91; Indels 0; Gaps 0;
41
42 Oy 134 GAGGCGCTGTATATACATGGCTAGGANNNNACACNCCCTGACNNNTGGANACTGCTACGACACAT 193
43 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
44 Db 6376 GTCGACATCGACACTGTAGCGCTGGCGCACAGACGAGACATCTCAATTAAGAAAAAAGAAAA 6435
45
46 Oy 134 GATATATGCGCTGCTGCGATGCGCTGCGACAGATCAACCTACTGCGAGACCAACNAG 253
47 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
48 Db 6436 GAAATATCGACGAAATTTAAACATATTTGGGAGAGAACTGTATTTTACAAATACCTAG 6495
49
50 Oy 254 GAGACTTNNAGACAGCATGATTANNACACACACTATTCCTTTTGGCAAGAAANA 313
51 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
52 Db 6496 TGTCTTCCGACGATGATGCTTCACTTCCCATTTATTAAGCTATTTTAAATCTTCA 6555
53
54 Oy 314 CTNTCTTTGANAATNTTT 332
55 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
56 Db 6556 GTGATGTTTGAATTTT 6574
57
58 RESULT 10
59 US-08-195-006-1
60 Sequence 1, Application US/08195006
61 Patent No. 6083688
62 GENERAL INFORMATION:
63 APPLICANT: Lanza, Francois
64 APPLICANT: Phillips, David R.
65 APPLICANT: Cazenave, Jean-Pierre
66 TITLE OF INVENTION: Platelet Glycoprotein V Gene and Uses
67 NUMBER OF SEQUENCES: 43
68 CORRESPONDENCE ADDRESS:
69 ADDRESSEE: Townsend and Townsend Kourile and Crew
70 STREET: 379 Lytton Avenue
71 CITY: Palo Alto
72 STATE: California

```



```

COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,006
FILING DATE: 10-FEB-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/089,455
FILING DATE: 09-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: DOV, Karen B.
REGISTRATION NUMBER: 25,664
REFERENCE/DOCKET NUMBER: 12418-28
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SHO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7452 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: intron
LOCATION: 1462..2419
FEATURE:
NAME/KEY: CDS
LOCATION: 2422..4101
FEATURE:
NAME/KEY: misc_signal
LOCATION: 68..76
OTHER INFORMATION: /function= "putative TPA responsive"
OTHER INFORMATION: element"
OTHER INFORMATION: /label= TRE
FEATURE:
NAME/KEY: misc_signal
LOCATION: 471..478
OTHER INFORMATION: /function= "Ets-1 cis-acting"
OTHER INFORMATION: sequence"
OTHER INFORMATION: /label= Ets-1
FEATURE:
NAME/KEY: misc_signal
LOCATION: 493..502
OTHER INFORMATION: /function= "Ets-1 cis-acting"
OTHER INFORMATION: sequence"
OTHER INFORMATION: /label= Ets-1
FEATURE:
NAME/KEY: repeat_region
LOCATION: 593..881
OTHER INFORMATION: /rpt_type= "other"
OTHER INFORMATION: /label= Alu
FEATURE:
NAME/KEY: misc_signal
LOCATION: 897..904
OTHER INFORMATION: /function= "Ets-1 cis-acting"
OTHER INFORMATION: sequence"
OTHER INFORMATION: /label= Ets-1
FEATURE:
NAME/KEY: misc_binding
LOCATION: 1142..1149
OTHER INFORMATION: /function= "Spi binding site"
OTHER INFORMATION: /standard_name= "Spi"
FEATURE:
NAME/KEY: misc_signal
LOCATION: 1178..1184

```

```

1  OTHER INFORMATION: /function= "Ets-1 cis-acting
2  OTHER INFORMATION: sequences"
3  OTHER INFORMATION: /label= Ets-1
4  FEATURE:
5  NAME/KEY: TATA_signal
6  LOCATION: 1199..1203
7  FEATURE:
8  NAME/KEY: TATA_signal
9  LOCATION: 1263..1269
10 FEATURE:
11 NAME/KEY: misc_binding
12 LOCATION: 1285..1289
13 OTHER INFORMATION: /function= "GATA-1 binding site"
14 FEATURE:
15 NAME/KEY: misc_binding
16 LOCATION: 1321..1326
17 OTHER INFORMATION: /function= "GATA-1 binding site"
18 FEATURE:
19 NAME/KEY: misc_signal
20 LOCATION: 1365..1372
21 OTHER INFORMATION: /function= "Ets-1 cis-acting
22 OTHER INFORMATION: sequences"
23 OTHER INFORMATION: /label= Ets-1]
24 FEATURE:
25 NAME/KEY: repeat_region
26 LOCATION: 6133..6440
27 OTHER INFORMATION: /rpt_type= "other"
28 OTHER INFORMATION: /label= Alu
29 FEATURE:
30 NAME/KEY: misc_signal
31 LOCATION: 5610..5615
32 OTHER INFORMATION: /standard_name= "polyadenylation
33 OTHER INFORMATION: signal sequence"
34 FEATURE:
35 NAME/KEY: misc_signal
36 LOCATION: 6966..6971
37 OTHER INFORMATION: /standard_name= "polyadenylation
38 OTHER INFORMATION: signal sequence"
39 FEATURE:
40 NAME/KEY: misc_signal
41 LOCATION: 7224..7229
42 OTHER INFORMATION: /standard_name= "polyadenylation
43 OTHER INFORMATION: signal sequence"
44 FEATURE:
45 NAME/KEY: misc_signal
46 LOCATION: 7358..7363
47 OTHER INFORMATION: /standard_name= "polyadenylation
48 OTHER INFORMATION: signal sequence"
49 FEATURE:
50 NAME/KEY: misc_feature
51 LOCATION: 1..7452
52 OTHER INFORMATION: /standard_name= "Nucleotide
53 OTHER INFORMATION: sequence containing the human GPV gene"
54 US-08-195-006-1
55
56 Query Match 15.3%, Score 53.4; DB 3; Length 7452;
57 Best Local Similarity 39.2%, Pred. No. 1.5e+02;
58 Matches 78; Conservative 30; Mismatches 91; Indels 0; Gaps 0;
59
60 Oy 134 GAGCGCTGTATATACCTGACCTGACGNNNNACGCTGACGNNNTGGANACCTGTCAGCAGCAT 153
61 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
62 Db 6376 GTGGCGACCTGACCTGACCTGAGCTGGAGGACAGAGAGAGACCTGTCATCTCAAAAAAAGAAAAA 6435
63 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
64 Oy 154 GATGATACGCTGCTGCGAATGCGCTGCGACAGATCAACCTGACGAGACAGACCAAG 253
65 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
66 Db 6436 GAAATTTGACGATTAATTTAAACTAATTTGGGAGAGAACTCTATTATTTTACAAATGCTAG 6495
67 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
68 Oy 254 GAGGCTTTNNAGACAGAGATGTATTTANNAAGCAACGATACCTTATTCGTAAGAGANA 313
69 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
70 Db 6496 TGTTCCTGCGACGTAAGACATGCTTCATCTTCCATTTATTTAGCATTTTAATCTTGA 6555
71 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
72 Oy 314 CTNTGCTTTGGANATNTTT 332
73 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

Db 6556 GTGATGTTTAGAATTTT 6574

RESULT 11

PCT-US94-07644A-1

Sequence 1, Application PC/TUS9407644A

GENERAL INFORMATION:

APPLICANT: COR Therapeutics, Inc.

TITLE OF INVENTION: Platelet Glycoprotein V Gene and Uses

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESS: Townsend and Townsend Kourie and Crew

STREET: 379 Lytton Avenue

CITY: Palo Alto

STATE: California

COUNTRY: US

ZIP: 94301

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/07644A

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Dow, Karen B.

REGISTRATION NUMBER: 29,684

REFERENCE/DOCKET NUMBER: 012418-003000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 326-2400

TELEFAX: (415) 326-2422

INFORMATION FOR SEQ. ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7452 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORGANISM SOURCE: Homo sapiens

FEATURE:

NAME/KEY: intron

LOCATION: 1462..2419

FEATURE:

NAME/KEY: CDS

LOCATION: 2422..4101

FEATURE:

NAME/KEY: misc\_signal

LOCATION: 68..76

OTHER INFORMATION: /function= "Putative TPA responsive"

OTHER INFORMATION: /label= TRE

FEATURE:

NAME/KEY: misc\_signal

LOCATION: 471..478

OTHER INFORMATION: /function= "Ets-1 cis-acting"

OTHER INFORMATION: /label= Ets-1

FEATURE:

NAME/KEY: misc\_signal

LOCATION: 493..502

OTHER INFORMATION: /function= "Ets-1 cis-acting"

OTHER INFORMATION: /label= Ets-1

FEATURE:

NAME/KEY: repeat\_region

LOCATION: 593..861

OTHER INFORMATION: /rpt\_type= "other"

OTHER INFORMATION: /label= Alu

NAME/KEY: misc\_signal

LOCATION: 897..904

OTHER INFORMATION: /function= "Ets-1 cis-acting"

OTHER INFORMATION: /label= Ets-1

FEATURE:

NAME/KEY: misc\_binding

LOCATION: 1142..1149

OTHER INFORMATION: /function= "Sp1 binding site"

OTHER INFORMATION: /standard\_name= "sp1"

FEATURE:

NAME/KEY: misc\_signal

LOCATION: 1178..1184

OTHER INFORMATION: /function= "Ets-1 cis-acting"

OTHER INFORMATION: /label= Ets-1

FEATURE:

NAME/KEY: TATA\_signal

LOCATION: 1199..1203

FEATURE:

NAME/KEY: TATA\_signal

LOCATION: 1263..1269

FEATURE:

NAME/KEY: misc\_binding

LOCATION: 1285..1289

OTHER INFORMATION: /function= "GATA-1 binding site"

FEATURE:

NAME/KEY: misc\_binding

LOCATION: 1321..1326

OTHER INFORMATION: /function= "GATA-1 binding site"

FEATURE:

NAME/KEY: misc\_signal

LOCATION: 1365..1372

OTHER INFORMATION: /function= "Ets-1 cis-acting"

OTHER INFORMATION: /label= Ets-1

FEATURE:

NAME/KEY: repeat\_region

LOCATION: 6133..6440

OTHER INFORMATION: /rpt\_type= "other"

OTHER INFORMATION: /label= Alu

FEATURE:

NAME/KEY: misc\_signal

LOCATION: 5610..5615

OTHER INFORMATION: /standard\_name= "Polyadenylation"

OTHER INFORMATION: /label= Polyadenylation

FEATURE:

NAME/KEY: misc\_signal

LOCATION: 6966..6971

OTHER INFORMATION: /standard\_name= "Polyadenylation"

OTHER INFORMATION: /label= Polyadenylation

FEATURE:

NAME/KEY: misc\_signal

LOCATION: 7224..7229

OTHER INFORMATION: /standard\_name= "Polyadenylation"

OTHER INFORMATION: /label= Polyadenylation

FEATURE:

NAME/KEY: misc\_signal

LOCATION: 7358..7363

OTHER INFORMATION: /standard\_name= "Polyadenylation"

OTHER INFORMATION: /label= Polyadenylation

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1..7452

OTHER INFORMATION: /standard\_name= "Nucleotide"

OTHER INFORMATION: /label= Nucleotide

PCT-US94-07644A-1

Query Match 15.3%; Score 53.4; DB 5; Length 7452;

Best Local Similarity 39.2%; Pred. No. 1.5e+02;

Matches 78; Conservative 30; Mismatches 91; Indels 0; Gaps 0;

QY 134 GAGGCMCTGATTCATGCTAGGNNNNNCCGCTGACNNNNNGAAGACCTGCTGACGACAT 193

DB 6376 GTGCCACCTGACCTAGCTGCTGCGGCGACGAGGACGACCTCCATCTCAAAAAAAGAAAAA 6435



Db 640 TGACTGACACACCACTACTACTACATATATACGCAATGTCAGATAGATAACCGAC 591  
QY 328 TTTTGTAGAG 337  
Db 580 TTAGTGACAG 571

## RESULT 14

US-09-068-140A-14

Sequence 14, Application US/09068140A

Patent No. 6281409

GENERAL INFORMATION:

APPLICANT: Mary Rose Woodhead, Mark Andrew Taylor

APPLICANT: and Rex Michael Brennan

TITLE OF INVENTION: Blackcurrant Promoters and Genes

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

FILING DATE:

APPLICATION NUMBER: US/09/068.140A

CLASSIFICATION:

PRIORITY APPLICATION DATA:

FILING DATE: NO. 6281409ember 4, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Dinner, Data L.

REGISTRATION NUMBER: 33, 680

REFERENCE/DOCKET NUMBER: C70237

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5017

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 5150 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Rides nigrum

STRAIN: Ben Alder

US-09-068-140A-14

Query Match

Best Local Similarity 15.0%; Score 52.4; DB 4; Length 5150;

Matches 101; Conservative 30; Mismatches 131; Indels 0; Gaps 0;

Db 89 TCCCTCTCTGCGGCTTTCACCTACCCCGCTTACGTCAGTCAGGCGCCAGCNCCTGTATACA 148

QY 3740 TCCCGCGTTAGGCGGCTTACCTTGCACATATGAGCCAGGAAGGCTTGTGGA 3799

Db 149 TCCCTAGGNNNNACNCTGACNNNTGANNACTCTNCAGCACATGATGATGCGTCNC 208

QY 3800 GGTGTCGCGGAGACTGCGACAGATGAGAAATTGTTAGTCATGCTGCTGCTGTC 3859

Db 209 TGGGATGNGCCTGNCACGATTCACCTGTCAGACCCANCAAGGAGACTTNNAGACA 268

QY 3860 TCTGTGAGCTTTGGCTGCGAGTGTAGTACACCTCAGTGTGTCAAATCAGTGTGCA 3919

QY 269 GAATGATATTANNACNACTACTTANCCCTTTTGCAANGANAACNTTCTTGANAAT 328  
Db 3920 GTGCGAGTGTAGTCCCTTTTAACTTTGCACTAAAAAATAAGATTCAGTGTCTAAT 3979  
QY 329 NTTGAGAGATTTCATAATAAAA 350  
Db 3980 TTTGAGATTACACCAATATAAAA 4001

## RESULT 15

US-08-370-975B-1

Sequence 1, Application US/08370975B

Patent No. 5622851

GENERAL INFORMATION:

APPLICANT: Maley, Frank

APPLICANT: Maley, Gladys F.

TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon, Hargrave, Devans &amp; Doyle

STREET: Clinton Square, P.O. Box 1051

CITY: Rochester

STATE: New York

COUNTRY: USA

ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

FILING DATE: 10-JAN-1995

APPLICATION NUMBER: US/08/370.975B

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Timian, Susan J.

REGISTRATION NUMBER: 34, 103

REFERENCE/DOCKET NUMBER: 20894/80

TELECOMMUNICATION INFORMATION:

TELEPHONE: (716)263-1600

TELEFAX: (716)263-1636

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 26764 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

POSITION IN GENOME:

CHROMOSOME/SEGMENT: 4q35

US-08-370-975B-1

Query Match

Best Local Similarity 15.0%; Score 52.4; DB 1; Length 26764;

Matches 77; Conservative 30; Mismatches 91; Indels 0; Gaps 0;

Db 197 TGATNGCGTCNCTGGGAATGNGCGCTGNCAGATCACTACCTGACGAGACCANCAAGGA 256

QY 26540 ACGGTGTGCTCTGCCCCCTTCTTTAATCTCAATTTAATTTAATAACATGCTCAGTA 26599

Db 257 GCTTNGAGACAGAAATGATTTANNACNACTACCTTATNCCTTTTGCAANGANAACNT 316

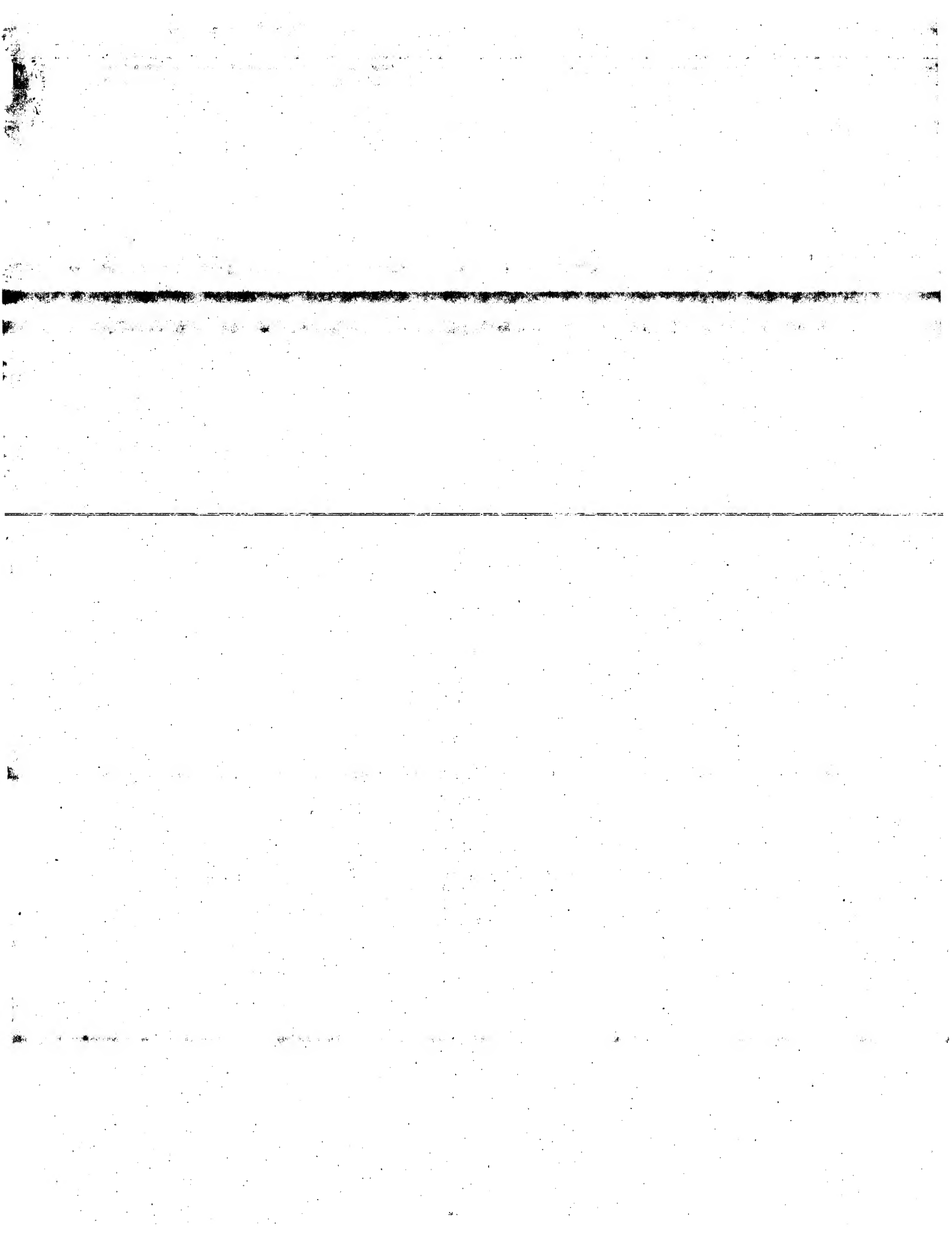
QY 26600 CCTGTGTGAGAAAGGCTTCTTTATCTTAAGATTTATACCTTTTAAAGTGTCTTA 26659

Db 317 TTTCTTGAANAATTTTGA 334

QY 26660 TATTTTCAATGAGTTTTTA 26677

Search completed: May 9, 2003, 20:41:29  
Job time : 99 secs

---





Db	3011	GCACGAGATGTTTCCAACTCTCCTGACCTCGTATGCCCATGGTGGCCCTCCCAAAATGCT	2952
Qy	140	CTGTATACATGCCCTAGGNNNNACNCCCTGACNNNTGGAAATCNCINACAGCATGATGA	199
Db	2951	GGGATTTACAGCGCTTAACCAACCATGCCCGGCTTGCACAAATATTTTCCACACCTATTTTCA	28922
Qy	200	TNCGCTCNCCTGGGAATGNGCCTTGCACAGATCAACCTACTGCAGACCANCNAAAGGAGCT	259
Db	2891	TCTTTTATTTTCTACGTCAACACGAGAGGTAGGCGACGAGCGAGATTTATTTATTTCA	28322
Qy	260	TNNGACACAAATGTATTANNAACNACATACCTTATNCTTTTGGAAANGAANAACCTNTTC	319
Db	2891	TTTATGCAACAATTTGTATATATATATATACATATTAACCATTTTCTGATGTCTCTGGAAAGTAACTTTGCT	2772
Qy	320	TTTGANAATNTT	332
Db	2771	TTTGATGATCTTAT	2759

RESULT 4  
US-09-764-877-3322/c  
Sequence 3322, Application US/0974877

RESULT 4  
 US-09-764-877-33322/C  
 Sequence 3322, Application US/09764877  
 Patent No. US20020147140A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antipoddes  
 FILE REFERENCE: PC005  
 CURRENT APPLICATION NUMBER: US/09/764,877  
 CURRENT FILING DATE: 2001-01-17  
 Prior application data removed - refer to PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 4031  
 SOFTWARE: Patentln Ver. 2.0  
 SEQ ID NO 3322  
 LENGTH: 8677  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-764-877-33322

Query Match                      16.3%; Score 57; DB 10; Length 8677;  
 Best Local Similarity    40.2%; Pred. No. 1.9e+02;  
 Matches    84; Conservative    30; Mismatches    95; Indels     0; Gaps     0;

121 CACTGATGCCAGAGGCNCTGTATACATGGCTTAGNNNNCACNCCTGACACNNNTGGANMC 180  
      ||||| . . . . . ||||| . . . . . : : : : . ||||| : |  
 Db 4977 CACTAGCGCCCTCACAAGAGCGTGTACATACAGGCCCAAGTCAGTGAGGACCTCTGGGGTTC 4918

RESULT 5  
 US-09-764-877-3321/c  
 ; Sequence 3321, Application US/09764877  
 ; Patent No. US20020147140A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```





SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 92

```

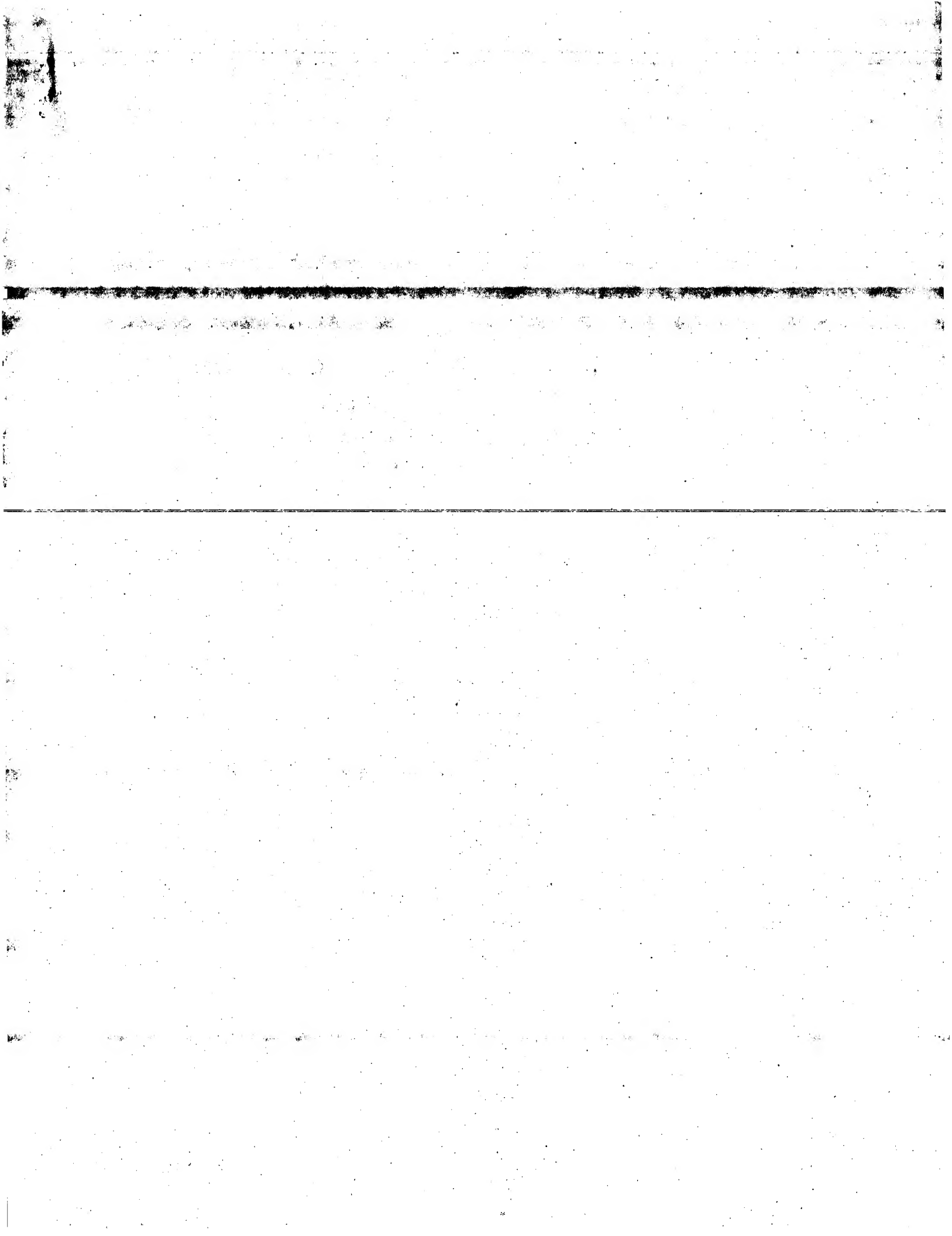
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-730-92

```

	Query Match	Similarity	Score	DB	Length
Best	Local	42.18	54.6	10	363
Matches	98	Conservative	30	Mismatches	104
				Indels	1
				Gaps	
QY	105	TTCCACTACCCCGCTTACTAGTCAGATGCCAGGCGNCGTATACCTGCTAGGNNNNACN	164		
DB	63	TTCCAGAAAGCGAGTTACTATGTTAGACTGGGATCTGTGTCACCAAGCTGCGTCACGGG	122		
QY	165	CCGTGACNNNTGGANACTNCTNCACACATGATGATGCGTCNCTGGGATGNCCTGNC	224		
DB	123	CCGTACCGACGAGCGCCTTTGTAGGGTTTAAAGAGGTAAAGG-AAAAGGCCAGGAGACATT	181		
QY	225	ACAGATCAACCTTACTGTCAGAACCANAAAGGAGCTTNNGAGACAGATGTAATTANNACC	284		

Db 182 TCTCAGCTTCATTCGCGCTTCAACCTGAGTGAACCTTGCCGAGAGAAATGTGCATATATGCC 241  
QY 285 NAGTACCTTATNGCTTTTGCAANGANAACTNTTCTTTGANANATNTTTGAGAG 337  
Db 242 AACGAGAAAATTGAGGATGCTGAGATTAAGCTCTATCTGTCTCTGGCTGGAGTG 294

Search completed: May 9, 2003, 20:47:52  
Job time : 402 secs



Run on: April 28, 2003, 15:31:39 ; Search time 35 Seconds

2665.012 million cell updates/sec

Title: US-09-880-711-328

Sequence: 1 atgaqctcacagtaactgca.....ttgaagatlttcataaaaa 350

Scoring table:

Xgapop	10.0	Xgapext	0.5
Ygapop	10.0	Ygapext	0.5
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Maximum DB Seq length: 20000000000

Post-processing: Minimum match 0%

### Listing first 45 summaries

Command line parameters:

```

O=/cgm2/1/USU01.spool/USF09880771/r1/matrix2804/2003.112137_833/4pp_query.fasta.1.513
DB-A.GeneSeq.101002.OPM1=fastan -supfx=92p_rag -minMATH=0.1 -LOOPTYPE=0
-LOOPTYPE=0 -UNIT=bits -START=1 -END=1 -MATH=1010sums2DX -IRVANS=human40.cdi
-LIST=45 -ALIGNEN=200 -THR SCORE=pct -THR MAX=100 -PIR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTTYPE=ptio -NORM=ext -HEATSIZE=500 -MINLEN=0 -MAXLEN=20000000000
-USUR=US09880771.@CGN1.1.114_@tina.2804/2003.112137_833 -MCPD=6 -ICPD=3
-NO_AMP -LACROBERT -NEG_SCORES=0 -WAIT -LONGLOC -DEFV -TCPO=3
-WARN=11HEU01=30 -THREDS=1 -XGAP=10 -XGAPEXT=0.5 -PCAP=6 -FCAPEXT=7
-PCAP=10 -YCAPEXT=0 -DELOP=6 -DELEXT=0.5 -FCAP=6 -FCAPEXT=7

```

Database :

1:	/SID52/qcdata/geneseq/geneseq-emb/AA1580.DAT *
2:	/SID52/qcdata/geneseq/geneseq-emb/AA1581.DAT *
3:	/SID52/qcdata/geneseq/geneseq-emb/AA1582.DAT *
4:	/SID52/qcdata/geneseq/geneseq-emb/AA1583.DAT *
5:	/SID52/qcdata/geneseq/geneseq-emb/AA1584.DAT *
6:	/SID52/qcdata/geneseq/geneseq-emb/AA1585.DAT *
7:	/SID52/qcdata/geneseq/geneseq-emb/AA1586.DAT *
8:	/SID52/qcdata/geneseq/geneseq-emb/AA1587.DAT *
9:	/SID52/qcdata/geneseq/geneseq-emb/AA1588.DAT *
10:	/SID52/qcdata/geneseq/geneseq-emb/AA1589.DAT *
11:	/SID52/qcdata/geneseq/geneseq-emb/AA1590.DAT *
12:	/SID52/qcdata/geneseq/geneseq-emb/AA1592.DAT *
13:	/SID52/qcdata/geneseq/geneseq-emb/AA1593.DAT *
14:	/SID52/qcdata/geneseq/geneseq-emb/AA1594.DAT *
15:	/SID52/qcdata/geneseq/geneseq-emb/AA1595.DAT *
16:	/SID52/qcdata/geneseq/geneseq-emb/AA1596.DAT *
17:	/SID52/qcdata/geneseq/geneseq-emb/AA1597.DAT *
18:	/SID52/qcdata/geneseq/geneseq-emb/AA1598.DAT *
19:	/SID52/qcdata/geneseq/geneseq-emb/AA1599.DAT *
20:	/SID52/qcdata/geneseq/geneseq-emb/AA1599.DAT *
21:	/SID52/qcdata/geneseq/geneseq-emb/AA2000.DAT *
22:	/SID52/qcdata/geneseq/geneseq-emb/AA2001.DAT *
23:	/SID52/qcdata/geneseq/geneseq-emb/AA2002.DAT *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	132	24.8	370	22	AAG64541	Genetic hydrophobin
2	132	24.6	158	22	AAU30653	Novel human secretin
3	131	24.6	730	18	AAW13888	Genetic formula #1
4	131	25.9	730	18	AAW13888	Genetic formula #1
5	130	25.7	119	22	AAU07680	Novel human dihydro
6	129	24.2	731	18	AAW13871	Genetic formula #2
7	129	25.5	731	18	AAW13871	Genetic formula #2
8	129	24.2	731	21	AAW13907	Genetic formula of
9	129	25.5	731	21	AAW13907	Genetic formula of
10	128	25.3	411	21	AAW77913	Plasmodium DBL gen
11	127	23.8	440	23	AAU021480	Consensus protein
12	126	24.9	740	19	AAW37150	Morfinic Menu, Ena-V
13	125.5	23.5	155	21	AAV49970	Hydrophobin generi
14	125	24.7	370	22	AAG64541	Genetic hydrophobi
15	124	23.3	176	21	AAV75385	EGF-like domain co
16	124	23.3	176	23	AAE17038	Human G protein-co
17	124	23.3	357	20	AAV37557	C. trachomatis proc
18	123	24.3	126	21	AAW48654	Zvegrit growth fact
19	123	24.3	134	22	AAU30506	Novel human secret
20	123	24.3	176	21	AAV79385	EGF-like domain co
21	123	24.3	176	23	AAE17038	Human G protein-co
22	123	23.1	261	21	AAV84325	Consensus sequence
23	123	24.3	328	14	AAK37616	Sequence of selecte
24	123	24.3	442	19	AAW86283	L. infantum cysteic
25	123	24.3	442	20	AAV43539	A. leishmania infan
26	122	24.1	116	21	AAV96859	Cysteine knot conse
27	122	24.1	176	23	AAU70945	Canine intercalin
28	122	24.1	375	21	AAV74405	Neisseria meningit
29	122	22.9	813	21	AAW01674	F1S2 protein sequen
30	121.5	22.8	112	23	AAU77106	Frog transforming
31	121.5	22.8	112	23	AAU77107	Frog transforming
32	121.5	22.8	112	23	AAU77108	Frog transforming
33	121	22.7	392	20	AAV38681	Neisseria meningit
34	121	23.9	192	20	AAV38681	Neisseria meningit
35	120	23.7	129	23	AAU76687	Eastern cottonmouth
36	120	22.5	134	22	AAU30506	Novel human secret
37	120	23.7	177	20	AAV37252	Amino acid sequenc
38	120	23.7	362	21	AAV77912	Plasmodium DBL gen
39	120	23.7	374	14	AAK33734	Human genetic sequ
40	120	23.7	411	21	AAV77914	Plasmodium DBL gen
41	120	22.5	443	20	AAV31670	Human IgG2 chain C
42	120	23.7	443	20	AAV31670	Human IgG2 chain C
43	120	22.5	444	20	AAV31672	Human IgG4 chain C
44	120	23.7	444	20	AAV31672	Human IgG4 chain C
45	120	22.5	447	20	AAV31669	Human IgG1 chain C

ALIGNMENTS

RESULT 1

AAG64541

ID AAG64541 standard; Protein; 370 AA.

AA64541:

18-OCF-2001 (first entry)

Generic hydrophobin protein purification.

Hydrophobin: protein purification.

Synthetic.

Key

Misc-difference

Location/Qualifiers

/label= Unknown

/note="X" = any amino acid, residue represents Xn, where n represents an integer"

FT

FT

```

FT Misc-difference 3..52
FT /label= unknown
FT /note= "X = any amino acid, optionally deleted for 49
FT amino acids"
FT
FT Misc-difference 54..58
FT /label= unknown
FT /note= "X = any amino acid, optionally deleted for 5
FT amino acids"
FT
FT Misc-difference 60..159
FT /label= unknown
FT /note= "X = any amino acid, optionally deleted for 99
FT amino acids"
FT
FT Misc-difference 161..260
FT /label= unknown
FT /note= "X = any amino acid, optionally deleted for 99
FT amino acids"
FT
FT Misc-difference 262..311
FT /label= unknown
FT /note= "X = any amino acid, optionally deleted for 49
FT amino acids"
FT
FT Misc-difference 313..317
FT /label= unknown
FT /note= "X = any amino acid, optionally deleted for 5
FT amino acids"
FT
FT Misc-difference 319..368
FT /label= unknown
FT /note= "X = any amino acid, optionally deleted for 49
FT amino acids"
FT
FT Misc-difference 370
FT /label= unknown
FT /note= "X = any amino acid, residue represents Xm, where
FT m represents an integer"
FT
FT WO200157076-A1.
FT
FT 09-AUG-2001.
FT
FT 02-FEB-2001; 2001WO-NL00083.
FT
FT 04-FEB-2000; 2000GB-0002661.
FT
FT (NANO-) APPLIED NANOSYSTEMS BV.
FT
FT De Vocht ML, Woesten HAB, Wessels JGH;
FT
FT WPI: 2001-522347/57.
FT
FT Purifying hydrophobin in hydrophobin containing solution by adsorbing
FT the solution to surface, contacting the surface with a surfactant
FT solution at specific temperature and separating desorbed hydrophobin
FT from surface
FT
FT
FT Disclosure: Page 2; 10pp: English.
FT
XX The invention relates to purifying hydrophobin from a hydrophobin
XX containing solution, comprising contacting the solution with a surface
XX for adsorption and separating the adsorbed hydrophobin from the solution.
XX Subsequently the surface is contacted with a solution containing a
XX surfactant at a temperature lower than 90°C. The present sequence is that
XX of a generic hydrophobin protein sequence of the invention.
XX
SQ Sequence 370 AA:

```

```

Alignment Scores:
Pred. No.: 1.08e-07 Length: 370
Score: 132.00 Matches: 2
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.77% Indels: 0
DB: 22 Gaps: 0

```

US-09-880-711-328 (1-350) x AAG64541 (1-370)

```

QY 138 CNCTGT 143
Db 1 ***Cys 2

```

RESULT 2  
AAU30663  
ID AAU30663 standard; Protein; 158 AA.  
XX  
AC AAU30663;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Novel human secreted protein #1154.  
XX  
KW Human; vaccination; gene therapy; nutritional supplement;  
KW stem cell proliferation; hematopoiesis; nerve tissue regeneration;  
KW immune suppression; immune stimulation; anti-inflammatory; leukemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200179449-A2.  
XX  
PD 25-OCT-2001.  
XX  
PF 16-APR-2001; 2001WO-US08656.  
XX  
PR 18-APR-2000; 2000US-0552929.  
PR 26-JAN-2001; 2001US-0770160.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YL, Liu C, Drmanac RT;  
XX  
DR WPI: 2001-611725/70.  
XX  
PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
PT vaccination, testing and therapy  
XX  
PS Claim 20; Page 325-326; 765pp: English.  
XX  
XX The invention relates to novel human secreted polypeptides. The  
CC polypeptides and antibodies to the polypeptides are useful for  
CC determining the presence of or predisposition to a disease associated  
CC with altered levels of polypeptide. The polypeptides are also useful for  
CC identifying agents (agonists and antagonists) that bind to them. Cells  
CC expressing the proteins are useful for identifying a therapeutic agent  
CC for use in treatment of a pathology related to aberrant expression or  
CC physiological interactions of the polypeptide. Vectors comprising  
CC the nucleic acids encoding the polypeptides and cells genetically  
CC engineered to express them are also useful for producing the proteins.  
CC The proteins are useful in genetic vaccination, testing and  
CC therapy, and can be used as nutritional supplements. They may be used to  
CC increase stem cell proliferation; to regulate haematopoiesis; and in  
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration.  
CC immune suppression and/or stimulation, as anti-inflammatory agents; and  
CC in treatment of leukemias. AAU29510-AAU3304 represent the amino acid  
XX sequences of novel human secreted proteins of the invention.

SQ Sequence 158 AA:

```

Alignment Scores:
Pred. No.: 1.49e-07 Length: 158
Score: 131.00 Matches: 9
Percent Similarity: 56.25% Conservative: 0
Best Local Similarity: 56.25% Mismatches: 7
Query Match: 24.58% Indels: 0
DB: 22 Gaps: 0

```

US-09-880-711-328 (1-350) x AAU30663 (1-158)

```

QY 26 CCCCACTAGCCCTGAGTGTGGCCCGCTACCTCTGTACCCAGCGGA 73
Db 25 ProAsnSerProAsnGlyGlnGlyProValAspSerProSerGly 40

```



RESULT 3  
AAM13888  
ID AAM13888 standard: protein; 730 AA.  
XX  
AC AAM13888:  
DT 13-MAY-1997 (first entry)  
XX  
DE Generic formula #1 of ant toxin of the invention.  
XX  
KW Toxin; ant; Bacillus thuringiensis; hymenopteran pest; pharaoh ant;  
XX biological control; Monomorium pharaonis; generic.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 2 /label= Ala, Thr  
FT Misc-difference 3 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 4 /label= Leu, Ile  
FT Misc-difference 5 /label= Asn, Gln  
FT Misc-difference 7 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 10 /note= "any naturally occurring amino acid except Cys, or  
FT absent"  
FT Misc-difference 11 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 12 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 14 /label= Asn, Gln  
FT Misc-difference 15 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 16 /label= Leu, Ile  
FT Misc-difference 17 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 18..24 /note= "any naturally occurring amino acid except Cys, or  
FT absent"  
FT Misc-difference 25..35 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 36 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 37..38 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 39 /note= "any naturally occurring amino acid except Cys, or  
FT absent"  
FT Misc-difference 40 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 41 /label= Glu, Asp  
FT Misc-difference 42..44 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 45 /label= Lys, Arg  
FT Misc-difference 46..48 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 49 /label= Lys, Arg  
FT Misc-difference 50..51 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 52..56 /note= "any naturally occurring amino acid except Cys, or  
FT absent"  
FT Misc-difference 57 /note= "any naturally occurring amino acid except Cys"

FT Misc-difference 58 /label= Lys, Arg, Glu, Asp  
FT Misc-difference 59..62 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 63 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 64..72 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 73 /label= Leu, Ile  
FT Misc-difference 74..82 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 83 /label= Leu, Ile  
FT Misc-difference 84 /label= Gly, Ser  
FT Misc-difference 85 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 86 /label= Leu, Ile  
FT Misc-difference 87 /label= Gly, Ser  
FT Misc-difference 88 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 89 /note= "any naturally occurring amino acid except Cys, or  
FT absent"  
FT Misc-difference 90 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 92..102 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 103..104 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 105..106 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 107 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 108..118 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 120 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 121..122 /note= "any naturally occurring amino acid except Cys, or  
FT absent"  
FT Misc-difference 123 /label= Leu, Ile  
FT Misc-difference 124 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 125..126 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 127 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 128..130 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 131 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 132..133 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 134..135 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 136..138 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 139 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 140..147 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 148 /label= Asn, Gln  
FT Misc-difference 149..150 /note= "any naturally occurring amino acid except Cys"  
FT Misc-difference 151 /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 152

FT /note="any naturally occurring amino acid except Cys"  
FT /label= Gly, Ser  
FT Misc-difference 153  
FT /label= Leu, Ile  
FT Misc-difference 154  
FT /label= Asn, Gln  
FT Misc-difference 155  
FT /note="any naturally occurring amino acid except Cys"  
FT /label= Ala, Thr  
FT Misc-difference 156  
FT /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 165  
FT /note="any naturally occurring amino acid except Cys"  
FT /note="any naturally occurring amino acid"  
FT Misc-difference 170  
FT /note="any naturally occurring amino acid"  
FT Misc-difference 171  
FT /note="any naturally occurring amino acid except Cys"  
FT /note="any naturally occurring amino acid except Cys, or  
FT /note="any naturally occurring amino acid except Cys, or  
FT Misc-difference 175  
FT /note="any naturally occurring amino acid except Cys"  
FT /note="any naturally occurring amino acid except Cys"  
FT Misc-difference 190  
FT /note="any naturally occurring amino acid except Cys"  
FT /note="any naturally occurring amino acid"  
FT Misc-difference 193  
FT /note="any naturally occurring amino acid"  
FT Misc-difference 194  
FT /note="any naturally occurring amino acid except Cys, or  
FT /note="any naturally occurring amino acid except Cys"  
FT Misc-difference 206  
FT /note="any naturally occurring amino acid except Cys"  
FT /label= Asn, Gln  
FT Misc-difference 208  
FT /label= Gly, Ser  
FT Misc-difference 209  
FT /note="any naturally occurring amino acid except Cys"  
FT /label= Ala, Thr  
FT Misc-difference 213  
FT /note="any naturally occurring amino acid except Cys"  
FT /label= Leu, Ile  
FT Misc-difference 215  
FT /note="any naturally occurring amino acid except Cys"  
FT /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 218  
FT /note="any naturally occurring amino acid except Cys, or  
FT /note="any naturally occurring amino acid except Cys, or  
FT Misc-difference 223  
FT /note="any naturally occurring amino acid except Cys"  
FT /label= Glu, Asp  
FT Misc-difference 224  
FT /note="any naturally occurring amino acid except Cys"  
FT /note="any naturally occurring amino acid except Cys"  
FT Misc-difference 225  
FT /note="any naturally occurring amino acid except Cys"  
FT /note="any naturally occurring amino acid except Cys"  
FT Misc-difference 230

## Alignment Scores:

Pred. No.: 1.37e-07 Length: 730  
Score: 131.00 Matches: 2  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 24.588 Indels: 0  
DB: 18 Gaps: 0

US-09-880-711-328 (1-350) x AAW13888 (1-730)

OY 24 TACCCC 29

Db 8 Typro 9

RESULT 4

AAW13888  
ID AAW13888 standard; protein; 730 AA.  
XX  
AC AAW13888;  
XX  
DE 13-MAY-1997 (first entry)  
XX  
DE Generic formula #1 of ant toxin of the invention.  
XX  
KW Toxin; ant; Bacillus thuringiensis; hymenopteran pest; pharaoh ant;  
KM biological control; Monomorium pharaonis; generic.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FH Misc-difference 2  
FT /label= Ala, Thr  
FT Misc-difference 3  
FT /note="any naturally occurring amino acid except Cys"  
FT Misc-difference 4  
FT /label= Leu, Ile  
FT Misc-difference 5  
FT /label= Asn, Gln  
FT Misc-difference 7  
FT /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 10  
FT /note="any naturally occurring amino acid except Cys, or  
FT /note="any naturally occurring amino acid except Cys"  
FT Misc-difference 11  
FT /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 12  
FT /note="any naturally occurring amino acid except Cys"  
FT Misc-difference 14  
FT /label= Asn, Gln  
FT Misc-difference 15  
FT /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 16  
FT /label= Leu, Ile  
FT Misc-difference 17  
FT /note="any naturally occurring amino acid except Cys"  
FT /note="any naturally occurring amino acid except Cys, or  
FT Misc-difference 18  
FT /note="any naturally occurring amino acid except Cys, or  
FT Misc-difference 25  
FT /note="any naturally occurring amino acid except Cys"  
FT Misc-difference 36  
FT /label= Met, Leu, Ile, Val, Phe  
FT Misc-difference 37  
FT /note="any naturally occurring amino acid except Cys"  
FT /note="any naturally occurring amino acid except Cys, or  
FT /note="any naturally occurring amino acid except Cys, or  
FT Misc-difference 40  
FT /note="any naturally occurring amino acid except Cys"  
FT /label= Glu, Asp  
FT Misc-difference 41  
FT /note="any naturally occurring amino acid except Cys"  
FT /note="any naturally occurring amino acid except Cys"  
FT Misc-difference 45  
FT /label= Lys, Arg  
FT Misc-difference 46  
FT /note="any naturally occurring amino acid except Cys"  
FT /label= Lys, Arg  
FT Misc-difference 49  
FT /note="any naturally occurring amino acid except Cys"  
FT /label= Lys, Arg  
FT Misc-difference 50  
FT /note="any naturally occurring amino acid except Cys"  
FT /note="any naturally occurring amino acid except Cys"  
FT Misc-difference 52  
FT /note="any naturally occurring amino acid except Cys, or  
FT /note="any naturally occurring amino acid except Cys, or  
FT Misc-difference 57  
FT /note="any naturally occurring amino acid except Cys"  
FT /label= Lys, Arg, Glu, Asp  
FT Misc-difference 58

```

FT Misc-difference 59..62
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 63
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 64..72
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 73
PT /label= Leu, Ile
FT Misc-difference 74..82
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 83
PT /label= Leu, Ile
FT Misc-difference 84
PT /label= Gly, Ser
PT Misc-difference 85
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 86
FT /label= Leu, Ile
FT Misc-difference 87
FT /label= Gly, Ser
FT Misc-difference 88
FT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 89
FT /note= "any naturally occurring amino acid except Cys, or
absent"
PT Misc-difference 90
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 92..102
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 103..104
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 105..106
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 107
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 108..118
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 120
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 121..122
FT /note= "any naturally occurring amino acid except Cys, or
absent"
PT Misc-difference 123
PT /label= Leu, Ile
FT Misc-difference 124
PT /label= Met, Leu, Ile, Val, Phe
PT Misc-difference 125..126
PT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 127
PT /label= Met, Leu, Ile, Val, Phe
PT Misc-difference 128..130
PT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 131
PT /label= Met, Leu, Ile, Val, Phe
PT Misc-difference 132..133
PT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 134..135
PT /label= Met, Leu, Ile, Val, Phe
PT Misc-difference 136..138
PT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 139
PT /label= Met, Leu, Ile, Val, Phe
PT Misc-difference 140..147
PT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 148
PT /label= Asn, Gln
FT Misc-difference 149..150
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 151
PT /label= Met, Leu, Ile, Val, Phe
PT Misc-difference 152
PT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 153

```

```

FT /label= Gly, Ser
FT Misc-difference 154
FT /label= Leu, Ile
PT Misc-difference 155
PT /label= Asn, Gln
FT Misc-difference 156..163
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 164
PT /label= Ala, Thr
FT Misc-difference 165
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 166..169
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 170
PT /note= "any naturally occurring amino acid"
FT Misc-difference 171..174
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 175..189
PT /note= "any naturally occurring amino acid except Cys, or
absent"
FT Misc-difference 190..192
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 193
PT /note= "any naturally occurring amino acid"
FT Misc-difference 194..205
PT /note= "any naturally occurring amino acid except Cys, or
absent"
PT Misc-difference 206..207
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 208
PT /label= Asn, Gln
FT Misc-difference 209
PT /label= Gly, Ser
PT Misc-difference 210..211
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 212
PT /label= Ala, Thr
FT Misc-difference 213..214
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 215
PT /label= Leu, Ile
FT Misc-difference 215..217
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 218
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 219..222
PT /note= "any naturally occurring amino acid except Cys, or
absent"
FT Misc-difference 223
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 224
PT /label= Glu, Asp
FT Misc-difference 225..229
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 230..238

```

Alignment Scores:

Pred. No.:	137e-07	Length:	730
Score:	131.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	25.89%	Indels:	0
DB:	18	Gaps:	0

US-09-880-711-328 (1-350) x AAM13888 (1-730)

OY 217 N:ATTCGACAGNC 206  
 |||::|11111  
 Db 7 \*\*\*tyrpro\*\*\* 10

RESULT 5  
 Abg07680  
 ID Abg07680 standard; protein: 119 AA.

XX ABG07680;  
AC 13-FEB-2002 (first entry)  
XX  
DT  
XX Novel human diagnostic protein #7671.  
DE  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS71867.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 20; SEQ ID No 38039; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 119 AA;

Alignment Scores:  
Pred. No.: 1,99e-07 Length: 119  
Score: 130.00 Matches: 3  
Percent Similarity: 80.00% Conservatve: 1  
Best Local Similarity: 60.00% Mismatches: 1  
Query Match: 25.69% Indels: 0  
DB: 22 Gaps: 0

US-09-880-711-328 (1-350) x ABG07680 (1-119)

OY 30 TGGGCTAAGTCGAG 16  
DB 110 TTPGlyLysAlaLys 114

RESULT 6  
AAW13871  
ID AAW13871 standard; protein; 731 AA.  
XX  
XX AAW13871;  
XX  
XX 13-MAY-1997 (first entry)  
XX  
XX  
DE Generic formula #2 of ant toxin of the invention.  
KW Toxin; ant; Bacillus thuringiensis; hymenopteran pest; pharaoh ant;  
KW biological control; Monomorium pharaonis; generic.  
XX  
XX Synthetic.  
XX  
XX  
XX Key Location/Qualifiers  
FH  
FT Misc-difference 2..7 /note= "unspecified amino acids"  
FT  
FT Misc-difference 10..12 /note= "unspecified amino acids"  
FT  
FT Misc-difference 14..90 /note= "unspecified amino acids"  
FT  
FT Misc-difference 92..118 /note= "unspecified amino acids"  
FT  
FT Misc-difference 120..229 /note= "unspecified amino acids"  
FT  
FT Misc-difference 231..240 /note= "unspecified amino acids"  
FT  
FT Misc-difference 242 /note= "unspecified amino acids"  
FT  
FT Misc-difference 244..249 /note= "unspecified amino acid"  
FT  
FT Misc-difference 251..319 /note= "unspecified amino acids"  
FT  
FT Misc-difference 321..335 /note= "unspecified amino acids"  
FT  
FT Misc-difference 336..342 /note= "unspecified amino acids"  
FT  
FT Misc-difference 343..472 /note= "unspecified amino acids"  
FT  
FT Misc-difference 474..541 /note= "unspecified amino acids"  
FT  
FT Misc-difference 543..637 /note= "unspecified amino acids"  
FT  
FT Misc-difference 638..644 /note= "unspecified amino acids"  
FT  
FT Misc-difference 646..686 /note= "unspecified amino acids"  
FT  
FT Misc-difference 688..728 /note= "unspecified amino acids"  
FT  
FT Misc-difference 730..731 /note= "unspecified amino acids"  
XX  
XX US5596071-A.  
PN  
XX 21-JAN-1997.  
PD  
XX  
XX  
XX 22-MAY-1991; 91US-0703977.  
PF  
XX  
XX 24-NOV-1993; 93US-0158232.  
PR  
XX 22-MAY-1991; 91US-0703977.  
PR  
XX 25-NOV-1991; 92US-0797645.  
PR  
XX 22-MAY-1992; 92US-0887980.  
XX  
XX (MYCO ) MYCOGEN CORP.  
PA  
XX Foncecerra L, Fu J, Kennedy MK, Meier H, Payne JM;  
PI Randall JB, Schnepf HE, Schwab GE, Uick HD;  
XX  
XX WPI; 1997-107615/10.  
DR  
XX  
XX Bacillus thuringiensis toxin - active against hymenopteran pests  
PT  
XX

PS Disclosure: Column 91-96: 64pp; English.  
XX  
CC AAW13888 and AAW13871 represent the generic formulae for the toxins of  
the invention. The toxins of the invention are derived from the  
CC Bacillus thuringiensis isolate PS8603 (NRRL B-18765), and are active  
CC against hymenopteran pests. The toxins can be used for the biological  
CC control of ants, particularly pharaoh ants (Monomorium pharaonis).  
XX  
SQ Sequence 731 AA:  
  
Alignment Scores:  
Pred. No.: 2.4e-07 Length: 731  
Score: 129.00 Matches: 2  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 24.20% Indels: 0  
DB: 18 Gaps: 0  
  
US-09-880-711-328 (1-350) x AAW13871 (1-731)  
  
QY 24 TACCCG 29  
|||||  
DB 8 TyrPro 9  
  
RESULT 7  
AAW13871  
ID AAW13871 standard; protein: 731 AA.  
XX  
AC AAW13871:  
XX  
DT 13-MAY-1997 (first entry)  
XX  
DE Generic formula #2 of ant toxin of the invention.  
XX  
DE  
XX  
KW Toxin; ant; Bacillus thuringiensis; hymenopteran pest; pharaoh ant;  
KW biological control; Monomorium pharaonis; generic.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 2..7  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 10..12  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 14..90  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 92..118  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 120..229  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 231..240  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 242  
FT Misc-difference /note= "unspecified amino acid"  
FT Misc-difference 244..249  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 251..319  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 321..335  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 336..342  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 343..472  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 474..541  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 543..637  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 638..644  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 646..686  
FT Misc-difference /note= "unspecified amino acids"  
FT Misc-difference 688..728

FT /note= "unspecified amino acids"  
FT Misc-difference 730..731  
FT /note= "unspecified amino acids"  
XX  
XX US5556071-A.  
XX  
XX 21-JAN-1997.  
XX  
XX  
XX 22-MAY-1991; 91US-0703977.  
XX  
XX  
XX 24-NOV-1993; 93US-0158232.  
XX  
XX 22-MAY-1991; 91US-0703977.  
XX 25-NOV-1991; 91US-0797645.  
XX 22-MAY-1992; 92US-0887980.  
XX  
XX (ATCO ) MYCOGEN CORP.  
XX  
XX Ponzerrada L, Po J, Kennedy MK, Meier H, Payne JM;  
XX Randall JH, Schnepf HR, Schwab GE, Ulick HJ;  
XX WPT; 1997-107915/10.  
XX  
XX  
XX Bacillus thuringiensis toxin - active against hymenopteran pests  
XX  
XX  
XX Disclosure: Column 91-96: 64pp; English.  
XX  
XX  
XX AAW13888 and AAW13871 represent the generic formulae for the toxins of  
the invention. The toxins of the invention are derived from the  
CC Bacillus thuringiensis isolate PS8603 (NRRL B-18765), and are active  
CC against hymenopteran pests. The toxins can be used for the biological  
CC control of ants, particularly pharaoh ants (Monomorium pharaonis).  
XX  
XX  
SQ Sequence 731 AA:  
  
Alignment Scores:  
Pred. No.: 2.4e-07 Length: 731  
Score: 129.00 Matches: 3  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 75.00% Mismatches: 0  
Query Match: 25.49% Indels: 0  
DB: 18 Gaps: 0  
  
US-09-880-711-328 (1-350) x AAW13871 (1-731)  
  
QY 217 NCATCCGACNG 206  
|||||  
DB 7 \*\*\*TyrPro\*\*\* 10  
  
RESULT 8  
AAW13907  
ID AAW13907 standard; Protein: 731 AA.  
XX  
AC AAW13907:  
XX  
XX 13-NOV-2000 (first entry)  
XX  
XX  
XX Generic formula of formulaid toxins.  
XX  
XX Hymenopteran; ant; pest control; 86Q3a; 17a; 17b; 33P2; 63B;  
KW formulaid.  
XX  
XX  
XX unidentified.  
XX  
XX  
XX Key Location/Qualifiers  
FT Misc-difference 1..731  
FT /note= "Xaa = any amino acid"  
XX  
XX US6077937-A.  
XX  
XX 20-JUN-2000.  
XX  
XX 16-OCT-1998; 98US-0173891.  
XX

PR 06-MAR-1996; 96US-0611928.  
 PR 24-NOV-1993; 93US-0158232.  
 PR 22-MAY-1991; 91US-0703977.  
 PR 25-NOV-1991; 91US-0797645.  
 PR 22-MAY-1992; 92US-0887980.  
 XX  
 PA (MYCO ) MYCOGEN CORP.  
 XX  
 PI Meier H, Kennedy MK, Schwab GE, Fu J, Payne JM, Uick HJ;  
 PI Foncerrada L, Schmepf HE, Randall JB;  
 XX  
 DR WPI; 2000-450980/39.  
 XX  
 PT New Bacillus thuringiensis toxins with activity against hymenopteran  
 PT pests such as fire ants and carpenter ants, conform to a specific  
 PT generic formula and have a specific amino acid sequence -  
 XX  
 PS Claim 1; Column 93; 67pp; English.  
 XX  
 CC The present invention relates to novel Bacillus thuringiensis toxins  
 CC with hymenopteran activity. Preparations containing protein from  
 CC Bacillus thuringiensis were tested for toxicity to ants. The N-terminal  
 CC amino acids of toxic proteins were then sequenced. These sequences were  
 CC used to design oligonucleotide probes. The probes were used to clone  
 CC anti-active toxin genes. The toxic proteins can be used to control  
 CC pests such as fire ants, carpenter ants, Argentine ants and pharaoh  
 CC ants. The proteins can also be used for producing transgenic plants  
 CC that are resistant to attack by ants. The proteins are a safe and  
 CC effective biological control agent against ant pests. The present  
 CC sequence is the generic formula of the formicidal toxins.  
 XX  
 SQ Sequence 731 AA;

Alignment Scores:  
 Pred. No.: 2.4e:07 Length: 731  
 Score: 129.00 Matches: 2  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 24.20% Indels: 0  
 DB: 21 Gaps: 0

US-09-880-711-328 (1-350) x AAB13907 (1-731)

OY 24 TACCCC 29

DB 8 TyrPro 9

RESULT 9

AAB13907  
ID AAB13907 standard; Protein; 731 AA.

XX AAB13907;

DT 13-NOV-2000 (first entry)

DE Generic formula of formicidal toxins.

XX Hymenopteran; ant; pest control; 86Q3a; 17a; 17b; 33F2; 63B;  
 KW formicidal.

XX Unidentified.

XX Key Location/Qualifiers

FT Misc-difference 1..731  
/note= "Xaa = any amino acid"

XX US6077937-A.

XX 20-JUN-2000.

XX 16-OCT-1998; 98US-0173891.

XX 06-MAR-1996; 96US-0611928.

PR 24-NOV-1993; 93US-0158232.  
 PR 22-MAY-1991; 91US-0703977.  
 PR 25-NOV-1991; 91US-0797645.  
 PR 22-MAY-1992; 92US-0887980.  
 XX  
 PA (MYCO ) MYCOGEN CORP.  
 XX  
 PI Meier H, Kennedy MK, Schwab GE, Fu J, Payne JM, Uick HJ;  
 PI Foncerrada L, Schmepf HE, Randall JB;  
 XX  
 DR WPI; 2000-450980/39.  
 XX  
 PT New Bacillus thuringiensis toxins with activity against hymenopteran  
 PT pests such as fire ants and carpenter ants, conform to a specific  
 PT generic formula and have a specific amino acid sequence -  
 XX  
 PS Claim 1; Column 93; 67pp; English.  
 XX  
 CC The present invention relates to novel Bacillus thuringiensis toxins  
 CC with hymenopteran activity. Preparations containing protein from  
 CC Bacillus thuringiensis were tested for toxicity to ants. The N-terminal  
 CC amino acids of toxic proteins were then sequenced. These sequences were  
 CC used to design oligonucleotide probes. The probes were used to clone  
 CC anti-active toxin genes. The toxic proteins can be used to control  
 CC pests such as fire ants, carpenter ants, Argentine ants and pharaoh  
 CC ants. The proteins can also be used for producing transgenic plants  
 CC that are resistant to attack by ants. The proteins are a safe and  
 CC effective biological control agent against ant pests. The present  
 CC sequence is the generic formula of the formicidal toxins.  
 XX  
 SQ Sequence 731 AA;

Alignment Scores:  
 Pred. No.: 2.4e:07 Length: 731  
 Score: 129.00 Matches: 3  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 75.00% Mismatches: 0  
 Query Match: 25.49% Indels: 0  
 DB: 21 Gaps: 0

US-09-880-711-328 (1-350) x AAB13907 (1-731)

OY 217 NCATGCCGANG 206

DB 7 \*\*\*TyrPro\*\*\* 10

RESULT 10

AAY77913  
ID AAY77913 standard; Protein; 411 AA.

XX AAY77913;

DT 13-JUN-2000 (first entry)

DE Plasmodium DBL gene family Proj3 conserved domain FL

XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;  
 KW DABP; Sialic Acid Binding Protein; SABB; malaria; vaccine; immunisation;  
 KW protozoacide; Proj3.

XX Plasmodium sp.

XX Key Location/Qualifiers

FT Misc-difference 1..411  
/note= "residues indicated Xaa are unspecified"

XX US593827-A.

XX 30-NOV-1999.

XX 07-JUN-1995; 95US-0487826.

XX 10-SEP-1993; 93US-0119677.

XX (USSH ) US EPPT HEALTH & HUMAN SERVICES.  
 PA  
 XX  
 P1 Sim KL, Chidnis C, Peterson DS, Su X, Wellens TE, Miller LH;  
 XX  
 DR WPI: 2000-194198/17.  
 XX  
 PT Isolated protein binding domains from Plasmodium vivax and Plasmodium  
 PT falciparum erythrocyte binding proteins useful for vaccinating against  
 PT malaria -  
 XX  
 PS Examples: Fig 1; 93pp; English.  
 XX  
 CC The invention relates to ebl-1 polypeptides that are encoded by the DBU  
 CC (duffy-binding like) gene family. The ebl-1 proteins are substantially  
 CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid  
 CC Binding Protein (SABP), which are soluble proteins that appear in the  
 CC culture supernatant after erythrocytes infected with malaria release  
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the  
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy  
 CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be  
 CC used to vaccinate against malaria, especially caused by P. falciparum.  
 CC Immunization with the polypeptide provides effective protection against  
 CC malaria.  
 CC  
 XX  
 SQ Sequence 411 AA:  
 Alignment Scores  
 Pred. No.: 3,266-07 Length: 411  
 Score: 128.00 Matches: 5  
 Percent Similarity: 88.69% Conservatave: 3  
 Best Local Similarity: 55.56% Mismatches: 1  
 Query Match: 25,30% Indels: 0  
 DB: Gaps: 0  
 US-09-880-711-328 (1-350) x AAV77913 (1-411)  
 QY 148 TGTATACAGNCCTCTGGCATGCAGT 122  
 DB 278 CYSIISeSCysLeuTYrAlaCysAsn 286  
 RESULT 11  
 ID AAO21480 standard: Protein: 440 AA.  
 AC AAO21480;  
 XX  
 DT 15-AUG-2002 (first entry)  
 DE  
 XX Consensus protein sequence for NGR's.  
 XX  
 CC Cerebroprotective; neuroprotective; cytosstatic; Nogo receptor homologue;  
 CC NGR3; axonal growth; central nervous system; CNS; cerebral injury;  
 CC spinal cord injury; stroke; demyelinating disease; multiple sclerosis;  
 CC monophasic demyelination; encephalomyelitis; Marchiafava-Bignami disease;  
 CC multifocal leukoencephalopathy; panencephalitis; Spongy degeneration;  
 CC Alexander's disease; Canavan's disease; metachromatic leukodystrophy;  
 CC Krabbe's disease; immune; bait protein; metachromatic leukodystrophy;  
 CC transgenic animal; unregulated cellular growth; cancer; tumour.  
 CC  
 XX  
 OS Unidentified.  
 XX  
 PN WO200229059-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PP 06-OCT-2001: 2001WO-US31488.  
 XX  
 PR 06-OCT-2000: 2000US-238361P.  
 XX  
 PA (UYVA ) UNIFY VALE.  
 PA (BIOJ ) BIOJEN INC.  
 XX

P1 Strilumatter SM, Cate RL, Sah DWY;  
 XX  
 DR WPI: 2002-415677/44.  
 XX  
 PT Novel Nogo receptor homology polypeptide, NGR2 or NGR3, useful for  
 PT treating central nervous system disorder, cerebral injury, spinal cord  
 PT injury, stroke, and demyelinating diseases -  
 XX  
 PS Example 16; Fig 1; 277pp; English.  
 XX  
 CC The invention relates to a Nogo receptor homologue polypeptide, NGR2 or  
 CC NGR3, comprising a 50 amino acid LRCT sequence, a 284 amino acid NTLRCT  
 CC sequence, or a 420, 461 or 392 amino acid sequence, all given in the  
 CC specification. The NGR3 protein or its binding antibody is useful for  
 CC decreasing inhibition of axonal growth of a central nervous system (CNS)  
 CC neuron, by contacting the neuron NGR3 or its antibody, and for treating  
 CC CNS disease, disorder or injury. NGR3 or a vector comprising NGR3 is  
 CC useful for treating cerebral injury, spinal cord injury, stroke,  
 CC demyelinating diseases, e.g. multiple sclerosis, monophasic  
 CC demyelination, encephalomyelitis, multifocal leukoencephalopathy,  
 CC panencephalitis, Marchiafava-Bignami disease, Spongy degeneration,  
 CC Alexander's disease, Canavan's disease, metachromatic leukodystrophy and  
 CC Krabbe's disease. NGR3 is useful for inducing an immune response in a  
 CC mammal against NGR3, as a bait protein in a two-hybrid or three-hybrid  
 CC assay, and as a research tool for identification, characterisation and  
 CC purification of interacting, regulatory proteins. The nucleotide  
 CC sequences of the invention are useful for screening for RFLP associated  
 CC with certain disorders, for genetic mapping, and for gene therapy. The  
 CC vector containing NGR3 is useful for producing non-human transgenic  
 CC animals. The NGR3 binding antibody is useful for isolating and purifying  
 CC NGR3, for localisation and/or quantitation of NGR3, and for diagnostic  
 CC and therapeutic purposes. The sequences of the invention, vectors and  
 CC antibodies are useful for treating or preventing unregulated cellular  
 CC growth such as cancer and tumour growth. This sequence represents the  
 CC consensus protein sequence for the NGR's of the invention.  
 CC  
 XX  
 SQ Sequence 440 AA:  
 Alignment Scores:  
 Pred. No.: 4,36-07 Length: 440  
 Score: 127.00 Matches: 3  
 Percent Similarity: 100.00% Conservatave: 1  
 Best Local Similarity: 75.00% Mismatches: 0  
 Query Match: 23.83% Indels: 0  
 DB: Gaps: 0  
 US-09-880-711-328 (1-350) x AAO21480 (1-440)  
 QY 10 CAGTACTGTCAC 21  
 DB 130 CIGTYrACTTYr 133  
 RESULT 12  
 ID AAW37150 standard: Protein: 740 AA.  
 AC AAW37150;  
 XX  
 DT 06-JUL-1998 (first entry)  
 DE  
 XX  
 DE Murine Menu, Ene-Vasp like (Evl) protein consensus.  
 XX  
 CC Ene-VASP like protein; Evi protein; Menu; mammalian Ene; Enabled;  
 CC cytoskeleton; cell morphology; cell adhesion; cell differentiation;  
 CC cell growth; cell motility; mouse.  
 CC  
 XX  
 OS Mus musculus.  
 XX  
 PN Key location/Qualifiers  
 PA Domain 1..113  
 PT /note= "Ene-VASP homology domain 1 (EVH1)"  
 PT Misc-difference 8  
 PT /note= "variable residue"

FT Misc-difference 12 /note="variable residue"  
FT Misc-difference 19..20 /note="variable residues"  
FT Misc-difference 26..29 /note="variable residues"  
FT Misc-difference 30..31 /note="variable residues"  
FT Misc-difference 37 /note="variable residue"  
FT Misc-difference 41..44 /note="variable residues"  
FT Misc-difference 56 /note="variable residue"  
FT Misc-difference 58 /note="variable residue"  
FT Misc-difference 65 /note="variable residue"  
FT Misc-difference 67 /note="variable residue"  
FT Misc-difference 94 /note="variable residue"  
FT Misc-difference 100..101 /note="variable residues"  
FT Misc-difference 104 /note="variable residues"  
FT Misc-difference 107..108 /note="variable residues"  
FT Misc-difference 112 /note="variable residue"  
FT Misc-difference 114..122 /note="variable residues"  
FT Misc-difference 124 /note="variable residue"  
FT Misc-difference 126..138 /note="variable residues"  
FT Misc-difference 140 /note="variable residue"  
FT Misc-difference 145 /note="variable residue"  
FT Misc-difference 148..150 /note="variable residues"  
FT Misc-difference 152 /note="variable residue"  
FT Misc-difference 155..239 /note="variable residues"  
FT Misc-difference 243 /note="variable residue"  
FT Misc-difference 245..413 /note="variable residues"  
FT Misc-difference 418..420 /note="variable residues"  
FT Misc-difference 422..423 /note="variable residues"  
FT Misc-difference 425..433 /note="variable residues"  
FT Misc-difference 438..444 /note="variable residues"  
FT Misc-difference 451..454 /note="variable residues"  
FT Misc-difference 460 /note="variable residues"  
FT Misc-difference 462..463 /note="variable residues"  
FT Misc-difference 466..468 /note="variable residues"  
FT Misc-difference 470..475 /note="variable residues"  
FT Misc-difference 477..478 /note="variable residues"  
FT Misc-difference 488..496 /note="variable residues"  
FT Misc-difference 498..513 /note="variable residues"

FT Misc-difference 519 /note="variable residues"  
FT Misc-difference 528..530 /note="variable residue"  
FT Misc-difference 532 /note="variable residues"  
FT Misc-difference 535..562 /note="variable residue"  
FT Misc-difference 565 /note="variable residues"  
FT Misc-difference 572..574 /note="variable residue"  
FT Misc-difference 580..592 /note="variable residues"  
FT Misc-difference 594..601 /note="variable residues"  
FT Misc-difference 603..620 /note="variable residues"  
FT Misc-difference 622..623 /note="variable residues"  
FT Misc-difference 625 /note="variable residues"  
FT Misc-difference 629..630 /note="variable residue"  
FT Misc-difference 633..637 /note="variable residues"  
FT Misc-difference 639..641 /note="variable residues"  
FT Misc-difference 643..702 /note="variable residues"  
FT Misc-difference 707 /note="variable residue"  
FT Misc-difference 713 /note="variable residue"  
FT Misc-difference 715..717 /note="variable residues"  
FT Misc-difference 720 /note="variable residue"  
FT Misc-difference 731 /note="variable residue"  
FT Misc-difference 735..740 /note="variable residues"  
FT Misc-difference 507..733 /note="C-terminal homology domain"  
PN WO9801755-A1.  
XX 15-JAN-1998.  
XX 03-JUL-1997; 97WO-US11669.  
XX 05-JUL-1996; 96US-0675815.  
XX (GBFB ) GENS BIOTECHNOLOGISCHE FORSCHUNG MBH.  
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
XX Gertler FB, Niebuhr K, Soriano P, Wehland J;  
XX WPT. 1998-101197/09.  
XX  
XX Detection of modulators of Mena and Ena-VASP-like genes and proteins  
XX - used in control of cytoskeletal dynamic events in normal and  
XX abnormal cell morphology, adhesion, motility, growth and  
XX differentiation  
XX  
XX Example 1: Page 68-70; 77pp; English.  
XX  
XX This polypeptide comprises a consensus of novel murine mammalian Ena  
XX (Mena) protein (see AA037148) and novel murine Ena-VASP like protein  
XX Evi (see AA037149). The greatest degree of amino acid identity is  
XX in the N-terminal 'Ena-VASP homology domain 1' that is responsible  
XX for Mena binding to Listeria, and to the cytoskeletal proteins





FT Misc-difference 370 amino acids"  
 FT /label= Unknown  
 FT /note= "X = any amino acid, residue represents Xm where  
 FT m represents an integer"  
 FT  
 FT WO200157076-A1.  
 FT  
 FT 09-AUG-2001.  
 FT  
 FT 02-FEB-2001; 2001WO-NL00083.  
 FT  
 FT 04-FEB-2000; 2000GB-0002661.  
 FT  
 FT (NANO-) APPLIED NANOSYSTEMS BV.  
 FT  
 FT De Vocht ML, Woesten HAB, Wessels JGH;  
 FT  
 FT WPI; 2001-522347/57.  
 FT  
 FT Purifying hydrophobin in hydrophobin containing solution by adsorbing  
 FT the solution to surface, contacting the surface with a surfactant  
 FT solution at specific temperature and separating described hydrophobin  
 FT from surface -  
 FT  
 FT PS Disclosure; Page 2; 10pp; English.  
 FT  
 FT CC The invention relates to purifying hydrophobin from a hydrophobin  
 FT containing solution, comprising contacting the solution with a surface  
 FT for adsorption and separating the adsorbed hydrophobin from the solution.  
 FT CC Subsequently the surface is contacted with a solution containing a  
 FT surfactant at a temperature lower than 900C. The present sequence is that  
 FT of a generic hydrophobin protein sequence of the invention.  
 FT  
 FT SO Sequence 370 AA:  
 FT  
 FT Alignment Scores:  
 FT Pred. No.: 7.57e-07 Length: 370  
 FT Score: 125.00 Matches: 3  
 FT Percent Similarity: 100.00% Conservative: 0  
 FT Best local Similarity: 100.00% Mismatches: 0  
 FT Query Match: 24.70% Indels: 0  
 FT DB: Gaps: 0  
 FT  
 FT US-09-880-711-328 (1-350) x AAG6541 (1-370)  
 FT  
 FT QY -190 TCGTGNAGN 182  
 FT  
 FT Db 2 Cys\*\*\*\*\* 4  
 FT  
 FT RESULT 15  
 FT ID AAY79385  
 FT AC AAY79385 standard; Protein; 176 AA.  
 FT  
 FT AC AAY79385;  
 FT  
 FT DT 01-AUG-2000 (first entry)  
 FT  
 FT XX EGF-like domain consensus.  
 FT  
 FT DE EGF-like domain consensus.  
 FT  
 FT XX SLGF; human G protein coupled receptor; signal transduction;  
 FT anti-proliferative; cardiant; hypotensive; antiarteriosclerotic;  
 FT antiarrhythmic; therapy; diagnosis; epidermal growth factor; EGF.  
 FT  
 FT OS Homo sapiens.  
 FT  
 FT FH Key Location/Qualifiers  
 FT FT Misc-difference 1..4  
 FT /note= "residues at positions 1-4 may be any amino  
 FT acid"  
 FT FT Misc-difference 5  
 FT /note= "conserved Cys residue involved in disulf"  
 FT FT Misc-difference 6..54

FT /note= "any 48 of the residues at positions 6-54  
 FT may be absent; intended a range of 0-48  
 FT amino acids"  
 FT  
 FT Disulfide-bond 55  
 FT /note= "conserved Cys residue involved in a  
 FT disulfide bond"  
 FT  
 FT Misc-difference 56..67  
 FT /note= "any 9 of the residues at positions 56-67  
 FT may be absent; intended a range of 3-12  
 FT amino acids"  
 FT  
 FT Disulfide-bond 68  
 FT /note= "conserved Cys residue involved in a  
 FT disulfide bond"  
 FT  
 FT Misc-difference 69..138  
 FT /note= "any 69 of the residues at positions 69-138  
 FT may be absent; intended a range of 1-70  
 FT amino acids"  
 FT  
 FT Disulfide-bond 139  
 FT /note= "conserved Cys residue involved in a  
 FT disulfide bond"  
 FT  
 FT Misc-difference 140..145  
 FT /note= "any 5 of the residues at positions 140-145  
 FT may be absent; intended a range of 1-6  
 FT amino acids"  
 FT  
 FT Disulfide-bond 146  
 FT /note= "conserved Cys residue involved in a  
 FT disulfide bond"  
 FT  
 FT Misc-difference 151..171  
 FT /note= "any 21 of the residues at positions 151-171  
 FT may be absent; intended a range of 0-21  
 FT amino acids"  
 FT  
 FT Misc-difference 173..174  
 FT /note= "residues at positions 173-174 may be any  
 FT amino acid"  
 FT  
 FT Disulfide-bond 175  
 FT /note= "conserved Cys residue involved in a  
 FT disulfide bond"  
 FT  
 FT Misc-difference 176  
 FT /note= "residue at position 176 may be any amino  
 FT acid"  
 FT  
 FT WO200018923-A2.  
 FT  
 FT PD 06-APR-2000.  
 FT  
 FT PF 30-SEP-1999; 99WO-US22923.  
 FT  
 FT PR 30-SEP-1998; 98US-0163821.  
 FT  
 FT PA (MTL-) MILLENNIUM PHARM INC.  
 FT  
 FT TSai F;  
 FT  
 FT WPI; 2000-293158/25.  
 FT  
 FT DR The present sequence is that of an epidermal growth factor (EGF)  
 FT like domain consensus. In an embodiment of the invention, a novel  
 FT human G protein coupled receptor, SLGF (see AAY79383), is identified  
 FT on the basis of the presence of an EGF-like domain in the protein.  
 FT The invention provides SLGF polynucleotides and polypeptides,  
 FT transgenic animals, fusion proteins, antigenic peptides,  
 FT antibodies, and methods for identifying modulator compounds.  
 FT Therapeutic and diagnostic methods utilizing these products are  
 FT also provided, especially for disorders associated with aberrant  
 FT SLGF expression such as proliferative, differentiative or

CC developmental, cardiac (hypertension, atherosclerosis coronary  
CC artery disease, arrhythmias), and haematopoietic disorders.

XX  
SQ Sequence 176 AA:

Alignment Scores:

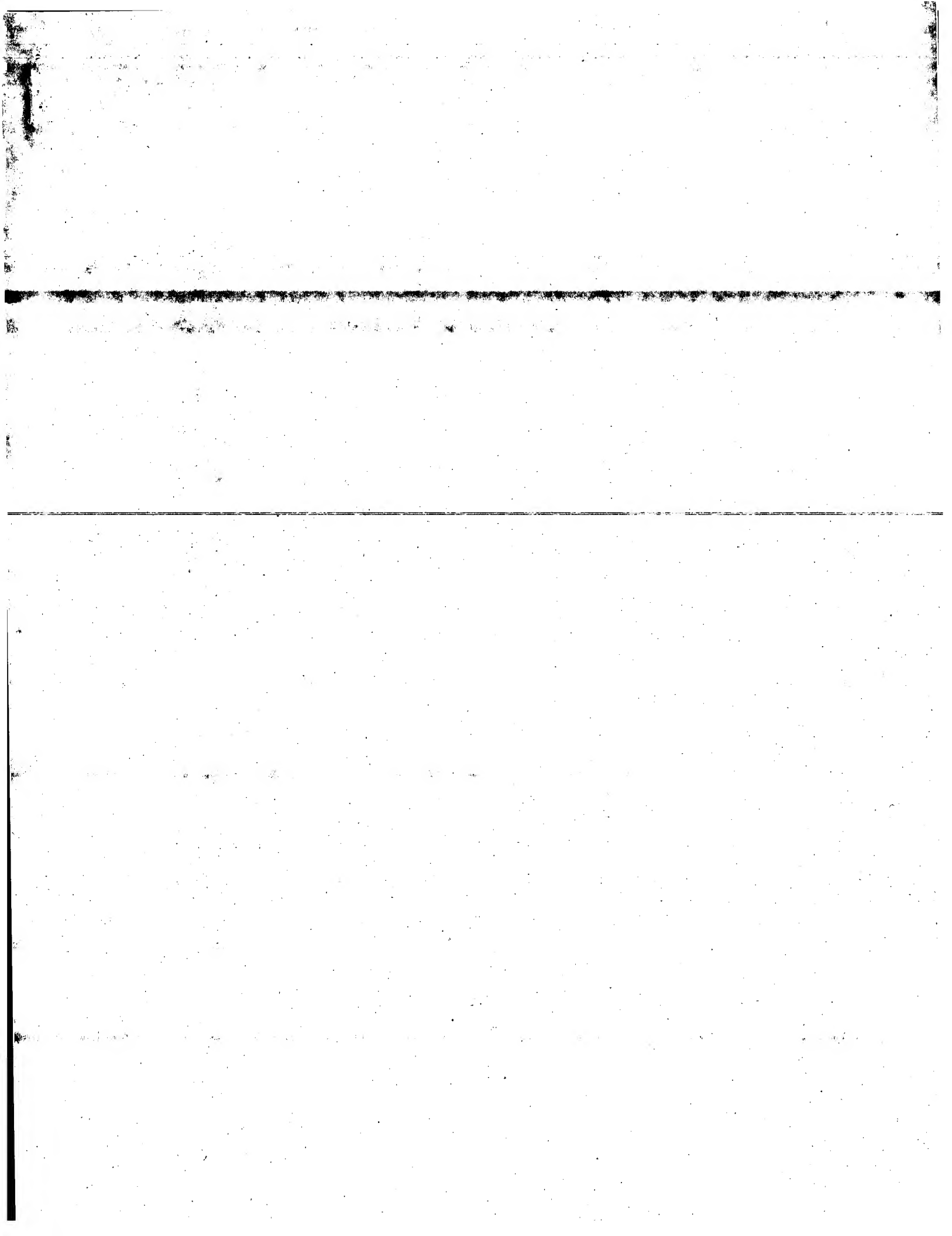
Pred. No.:	1.04e-06	Length:	176
Score:	124.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	23.268	Indels:	0
DH:	21	Gaps:	0

US-09-880-711-32E (1-350) x AAY79385 (1-176)

OY 254 GGAGCTTNN 262

DH 149 G1YAla\*\*\* 151

Search completed: April 28, 2003, 15:36:04  
Job time : 39 secs



GenCore version 5.1.4.F5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_n2p model

Run on: April 28, 2003, 15:34:44 : Search time 15.5 Seconds

(without alignments)  
3450.979 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 533

Sequence: 1 atgagccacgctacccgcga.....ttgcgagatctcaataaaaa 350

Scoring table:

BL0SUM62DX	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Dgapop 6.0 , Dextext 7.0	

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame_n2p.model -DEV=xlp  
-O=/con2_1/USPTC-SPOOL/US09880711/runat_28042003_112139_867/app_query.fasta_1_519  
-DB=PIR_73 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -DOOPCL=0 -DOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62dx -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=PCL -THR.MAX=100 -THR.MIN=3 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pl0 -NOM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09880711.acgn.1.1.62.grunal.28042003_112139_867 -NCP=6 -ICPU=3  
-NO_XLPHYX -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOC -DEV.TIMEOUT=120  
-NARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAEXT=0.5 -DELAP=6 -DELEXT=7
```

Database :  
1: PIR\_73:\*  
2: pir1:\*  
3: pir2:\*  
4: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	225.5	42.3	196	JC6127	RNA-binding protei
2	85.5	16.0	213	T13752	NADH2 dehydrogenas
3	79	15.6	213	T13752	NADH2 dehydrogenas
4	78.5	14.7	1870	S37671	MHC class III hist
5	78.5	14.7	1872	S36152	MHC class III hist
6	78.5	14.7	2142	B35098	MHC class III hist
7	75	14.1	212	T13571	NADH2 dehydrogenas
8	75	14.1	217	T13660	NADH2 dehydrogenas
9	74	13.9	228	S53504	unknown protein [l
10	74	13.7	652	T12587	extensin-like prot
11	73	13.5	579	E75275	NADH2 dehydrogenas
12	72	13.5	579	E75275	hypothetical prote
13	72	13.5	1159	T43461	probable phosphodi
14	70	13.1	494	S39607	transcription fact

15	70	13.1	542	2	S39608	transcription fact
16	70	13.8	578	2	T13664	NADH2 dehydrogenas
17	70	13.1	631	1	A36749	transcription fact
18	69.5	13.7	217	2	T13660	NADH2 dehydrogenas
19	69	12.9	427	2	G70555	hypothetical prote
20	68.5	12.9	427	2	E87669	conserved hypothet
21	68.5	12.9	1495	2	S60255	transcription co-r
22	68	12.8	428	2	T24769	hypothetical prote
23	68	13.4	652	2	T12587	NADH2 dehydrogenas
24	66.5	12.5	744	2	E86255	hypothetical prote
25	66.5	12.5	873	2	A47283	calphotin - fruit
26	66	14.0	212	2	T13571	NADH2 dehydrogenas
27	66	12.4	407	2	C70815	hypothetical prote
28	66	12.4	578	2	T13664	NADH2 dehydrogenas
29	65.5	12.3	548	2	A57640	retinoblastoma bin
30	65	12.2	134	2	D84672	hypothetical prote
31	65	12.2	228	2	S46965	microfilament shea
32	65	12.2	356	1	WJH02H	homeotic protein H
33	65	12.2	519	2	T45764	hypothetical prote
34	65	12.8	544	2	S76859	hypothetical prote
35	64.5	12.7	232	2	JC7705	hypothetical prote
36	64.5	12.1	395	2	H75457	beta crystallin B1
37	64.5	12.1	751	2	S68957	hypothetical prote
38	64.5	12.1	1445	2	T50508	adhesive plaque pr
39	64.5	12.1	1445	2	S68957	hypothetical prote
40	64	12.6	207	2	T12662	NADH2 dehydrogenas
41	64	12.0	212	2	T13367	microfilament shea
42	64	12.0	237	2	S46964	regulatory protein
43	64	12.0	432	2	A28913	hypothetical prote
44	64	12.0	447	2	T49350	hypothetical prote
45	63.5	11.9	233	2	T47136	hypothetical prote

#### ALIGNMENTS

```
RESULT 1  
JC6127  
RNA-binding protein type 1 - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000  
C:Accession: JC6127  
R:Shimamoto, A.; Kikuo, S.; Ichikawa, K.; Suzuki, N.; Yamabe, Y.; Imanura, O.; Tokuta  
Futuhili, Y.  
Proc. Natl. Acad. Sci. U.S.A. 93, 10513-10517, 1996  
A:Title: A unique human gene that spans over 230 kb in the human chromosome 8p11-12 a  
A:Reference number: JC6127; MIMID:97008106; PMID:8855282  
A:Accession: JC6127  
A:Molecule type: mRNA  
A:Residues: 1-196 <SH1>  
A:Cross-references: DDBJ:D84107; NID:q1669546; PIDD:BA12225.1; PID:q1669547  
C:Comment: This protein plays a role in RNA metabolism.  
C:Genetics:  
A:Gene: GDB:RBPM5  
A:Cross-references: GDB:9956209; OMIM:601558  
A:Map position: 8p12-8p11  
A:Superfamily: human RNA-binding protein type 1; ribonucleoprotein repeat homology  
C:Keywords: RNA binding  
F:25-88/Domain: ribonucleoprotein repeat homology <RKM2>
```

Alignment Scores: 3.68e-18 Length: 196  
Pred. No.: 225.50 Matches: 46  
Score: 225.50 Conservative: 6  
Percent Similarity: 83.87 Mismatches: 9  
Best Local Similarity: 76.19 Indels: 1  
Query Match: 42.318 Gaps: 1

US-09-880-711-328 (1-350) x JC6127 (1-196)

```
OY 3 GAGCTACAGTACCTGACTTTACCCAGTACGCTTGAAGTGTGGCCCGGCTACCTGCG 62  
DB 134 GlnAethrValProAlaLeuTyfProSerSerProGluValTrrPalatProTyfProPan 153
```

[illegible]

RESULT 2  
T13752

A:Name: ndhF  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5  
Keywords: chloroplast; membrane-associated complex; NAD: oxidoreductase  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 03-Jun-2002  
C:Accession: J13752  
R:Catalan, P.; Kelllogg, E.A.; Olmstead, R.G.  
Mol. Phylogenet. Evol. 8, 150-166, 1997  
A:Title: Phylogeny of Poaceae subfamily Pooideae based on chloroplast ndhF gene sequence  
A:Reference number: Z13533; M01D:97446273; PMID:9259221  
A:Accession: J13752  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-213 <CAT>  
A:Cross-references: EMBL:U71026; NID:q1763190; PID:q1763191; PIDN:AA836692.1  
A:Genetics:  
A:Genome: chloroplast  
A:Note: ndhF  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5  
Keywords: chloroplast; membrane-associated complex; NAD: oxidoreductase

Alignment Scores:

Pred. NO.:	0.0662	Length:	213
Score:	85.50	Matches:	9
Percent Similarity:	54.558	Conservative:	21
Best Local Similarity:	16.368	Mismatches:	13
Query Match:	16.048	Indels:	12
DB:	2	Gaps:	2

US-09-880-711-328 (1-350) x T13752 (1-213)

**OY**

157 ::: : :	NNNNCACMCCTGCANNNTGGANACTNCTNCACACATGATGATTGCCGTCCGTCGGCAATG	216
----------------------	--	-----

202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

Db  
44 PheSerGlnAsnIleProLysIleGln\*\*\*\*\*ThrArgAsnArgIleGlySerPhe\*\*\*\* 63

0Y 244 ACCAANCNAGGAGCGCTTNNAGACAGATATATTANNAACCACT 288  
 |||:::||||::: |||::: |||  
 Db 64 ThSerPheGyAlaIaLysLysThrPheAlaTyrProHISgluThr 78

### RESULT 3

NAMH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Triticum boeoticum chloroplast (T13752)  
 C:Species: chloroplast Triticum boeoticum  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 03-Jun-2002  
 C:Accession: T13752  
 R:Catalogan, P.; Kelllogg, E.A.; Olmstead, R.G.  
 Mol. Phylogenet. Evol. 8, 150-166, 1997  
 A:Title: Phylogeny of Poaceae subfamily Pooideae based on chloroplast ndhF gene sequencin  
 A:Reference number: Z15553; MIMD:97446273; PMID:9239221  
 A:Accession: T13752

A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA

A;Residues: 1-213 <CAT>  
A;Cross-references: EMBL:U71026; NTD:q1763190; PID:q1763191; PIDN:AAB3692.1

C;Genetics:  
A;Genome; chloroplast

A; note: name

[illegible]

US-09-880-711-328 (1-350) X T13752 (1-213)

OY 211 CCAGNGACGCNATCATTCATGCTGTCTGAGAGAGTNTCCANNCTCAGGNGTGNNMCC 155  
|||::: ::::: |||::: |||::: |||::: |||::: |||:::  
Db 25 ProLysylGlyAlaAsnArgAspPheValLeuSerThr\*\*LysSerGlyValSerPhe 43

## RESULT 4

S37671  
 MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu  
 C/Species: Homo sapiens (man)  
 C/Date: 20-Feb-1995 #sequence-revision 20-Feb-1995 #text-change 15-Sep-2000  
 C/Accession: S37671  
 R/Bouquelerec, L.  
 Submitted to the EMBL Data Library, August 1992  
 A/Reference number: S37671  
 A/Accession: S37671  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1870 <BOU>  
 A/Cross-references: EMBL:Z15025; NID:g29374; PID:g29375  
 C/Genetics:  
 A/Map position: 6p21.3  
 A/Intons: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;  
 A/Superfamily: collagen alpha 1(IV) chain

Alignment Scores:

Pred. No.:	0.286	Length:	1870
Score:	78.50	Matches:	30
Percent Similarity:	52.88	Conservative:	25
Best Local Similarity:	28.85	Mismatches:	37
Query Match:	14.73	Indels:	12
GB:	2	Gaps:	3

US-09-880-711-328 (1-350) x S37671 (1-1870)

15 CCTGCACCTTACCCCGAGTACCCCTGAAGTGTGGCCCGCTACCTCTGTACCCAGCGAG 74

D<sub>b</sub> 507 ProLaAlaProProLaAlaAlaProSerThrProAlaProProProLaValProLysGlu 52

QY 75 TTA---GCGCGCTCTTCTCCTCTCTGCGGCTTACCTACCCGGTTACATGCATGCC 13

132 CAGAGCCNCTGCTATACATGCCGTACCAGCCTTGGTTTCAG, -17  
 UU 321 LeuproLalProProLalProProLalaser-AlaProThrProglutylsuiProGI 54

[illegible]

178 -----NACTTCTNCAGACATGNATGATNCGCTNCTGGGAATGNGCCTGNC 22

Db 566 cThrLeuValSerGlyGlySerThrSerSerThrSerSerGlySerPheGluAla 58

225 ACAGATCAAC---CTACTGCAGACCANCAAGGAGCTTNNAGACAGATGTATTANNA 28

```

586 rProValGlnProGlnLeuProSerGlyGlyProGlnProGlnGlnValProPr 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      || || : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY 282 ACCNACTACC 291

Db 606 opProThrThr 609

## RESULT 5

MMHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu



Alignment Scores:	
Pred. No.:	1.09
Score:	75.00
Percent Similarity:	57.89%
Best Local Similarity:	19.30%
Query Match:	14.07%
Db:	2
Length:	217
Matches:	11
Conservative:	22
Mismatch:	10
Indels:	14
Gaps:	2

09	NACCTGCTGAGCAACATGATGATNGCGTGGGATGCTGCGACAGATCAACCTA	137
178	:     :     :     :     :     :     :     :     :     :	
Db	3 GlnSerTyrSerSerThrLysSerSer-----LeuTyrSerIleSerLeu	18

0y CTGCAGACCANCNAGGAGAGCTTNNAGACAGCAATGATTANNACCACTACC-----291  
 238 ::::::::::::::::::::|||  
 Db 19 Tptclglyargtllepolyscllyvalasnargaspheleuleuserthrlysser 38

QY	292	-----TTATNCCTTTGCACANGAAGCAACNTT	318
Db	39	GLYAlserPhepneSerGlnasnIleProtylsteIleGlnGly**ThnIar9	55

```

RESULT 9
686441
unknown protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
Accession: G66441

```

Rhede, L. G., 1990, *Journal of Polymer Science: Part A: Polymer Chemistry*, 28, 1565-1574.  
 Rhede, L. G., 1991, *Macromolecules*, 24, 1565-1574.  
 Rhede, L. G., 1992, *Macromolecules*, 25, 1565-1574.  
 Rhede, L. G., 1993, *Macromolecules*, 26, 1565-1574.  
 Rhede, L. G., 1994, *Macromolecules*, 27, 1565-1574.  
 Rhede, L. G., 1995, *Macromolecules*, 28, 1565-1574.  
 Rhede, L. G., 1996, *Macromolecules*, 29, 1565-1574.  
 Rhede, L. G., 1997, *Macromolecules*, 30, 1565-1574.  
 Rhede, L. G., 1998, *Macromolecules*, 31, 1565-1574.  
 Rhede, L. G., 1999, *Macromolecules*, 32, 1565-1574.  
 Rhede, L. G., 2000, *Macromolecules*, 33, 1565-1574.  
 Rhede, L. G., 2001, *Macromolecules*, 34, 1565-1574.  
 Rhede, L. G., 2002, *Macromolecules*, 35, 1565-1574.  
 Rhede, L. G., 2003, *Macromolecules*, 36, 1565-1574.  
 Rhede, L. G., 2004, *Macromolecules*, 37, 1565-1574.  
 Rhede, L. G., 2005, *Macromolecules*, 38, 1565-1574.  
 Rhede, L. G., 2006, *Macromolecules*, 39, 1565-1574.  
 Rhede, L. G., 2007, *Macromolecules*, 40, 1565-1574.  
 Rhede, L. G., 2008, *Macromolecules*, 41, 1565-1574.  
 Rhede, L. G., 2009, *Macromolecules*, 42, 1565-1574.  
 Rhede, L. G., 2010, *Macromolecules*, 43, 1565-1574.  
 Rhede, L. G., 2011, *Macromolecules*, 44, 1565-1574.  
 Rhede, L. G., 2012, *Macromolecules*, 45, 1565-1574.  
 Rhede, L. G., 2013, *Macromolecules*, 46, 1565-1574.  
 Rhede, L. G., 2014, *Macromolecules*, 47, 1565-1574.  
 Rhede, L. G., 2015, *Macromolecules*, 48, 1565-1574.  
 Rhede, L. G., 2016, *Macromolecules*, 49, 1565-1574.  
 Rhede, L. G., 2017, *Macromolecules*, 50, 1565-1574.  
 Rhede, L. G., 2018, *Macromolecules*, 51, 1565-1574.  
 Rhede, L. G., 2019, *Macromolecules*, 52, 1565-1574.  
 Rhede, L. G., 2020, *Macromolecules*, 53, 1565-1574.  
 Rhede, L. G., 2021, *Macromolecules*, 54, 1565-1574.  
 Rhede, L. G., 2022, *Macromolecules*, 55, 1565-1574.  
 Rhede, L. G., 2023, *Macromolecules*, 56, 1565-1574.  
 Rhede, L. G., 2024, *Macromolecules*, 57, 1565-1574.  
 Rhede, L. G., 2025, *Macromolecules*, 58, 1565-1574.

C.A.: Li, J.H.; Li, X.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maeli, R.; Mardatal, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, C.A.; Li, J.H.; Li, X.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maeli, R.; Mardatal, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G86441  
A:Status: preliminary

A:Residue: 1:1201 <STO>  
A:Cross-references: GB:AE005172; NID:g11136725; PIDN:AAG31306.1; GSPDB:GN00141  
Genetics:

Alignment Scores:	
Pred. No.:	0.793
Score:	75.00
Percent Similarity:	50.468
Length:	1201
Matches:	31
Conservative:	24

US-09-880-711-328 (1-350) x G86441 (1-1201

15 CCGACCTTTACCCCGATACCCCTGAAGTGTGGCCCCCGATCCCTGTACCCAGCGAG 74

[illegible]

7232 7233 7234 7235 7236 7237 7238 7239 7240 7241 7242 7243 7244 7245 7246 7247 7248 7249 7250 7251 7252 7253 7254 7255 7256 7257 7258 7259 7260 7261 7262 7263 7264 7265 7266 7267 7268 7269 7270 7271 7272 7273 7274 7275 7276 7277 7278 7279 7280 7281 7282 7283 7284 7285 7286 7287 7288 7289 7290 7291 7292 7293 7294 7295 7296 7297 7298 7299 7300 7301 7302 7303 7304 7305 7306 7307 7308 7309 7310 7311 7312 7313 7314 7315 7316 7317 7318 7319 7320 7321 7322 7323 7324 7325 7326 7327 7328 7329 7330 7331 7332 7333 7334 7335 7336 7337 7338 7339 7340 7341 7342 7343 7344 7345 7346 7347 7348 7349 7350 7351 7352 7353 7354 7355 7356 7357 7358 7359 7360 7361 7362 7363 7364 7365 7366 7367 7368 7369 7370 7371 7372 7373 7374 7375 7376 7377 7378 7379 7380 7381 7382 7383 7384 7385 7386 7387 7388 7389 7390 7391 7392 7393 7394 7395 7396 7397 7398 7399 7400 7401 7402 7403 7404 7405 7406 7407 7408 7409 7410 7411 7412 7413 7414 7415 7416 7417 7418 7419 7420 7421 7422 7423 7424 7425 7426 7427 7428 7429 7430 7431 7432 7433 7434 7435 7436 7437 7438 7439 7440 7441 7442 7443 7444 7445 7446 7447 7448 7449 7450 7451 7452 7453 7454 7455 7456 7457 7458 7459 7460 7461 7462 7463 7464 7465 7466 7467 7468 7469 7470 7471 7472 7473 7474 7475 7476 7477 7478 7479 7480 7481 7482 7483 7484 7485 7486 7487 7488 7489 7490 7491 7492 7493 7494 7495 7496 7497 7498 7499 7500 7501 7502 7503 7504 7505 7506 7507 7508 7509 7510 7511 7512 7513 7514 7515 7516 7517 7518 7519 7520 7521 7522 7523 7524 7525 7526 7527 7528 7529 7530 7531 7532 7533 7534 7535 7536 7537 7538 7539 7540 7541 7542 7543 7544 7545 7546 7547 7548 7549 7550 7551 7552 7553 7554 7555 7556 7557 7558 7559 7560 7561 7562 7563 7564 7565 7566 7567 7568 7569 7570 7571 7572 7573 7574 7575 7576 7577 7578 7579 7580 7581 7582 7583 7584 7585 7586 7587 7588 7589 7590 7591 7592 7593 7594 7595 7596 7597 7598 7599 7600 7601 7602 7603 7604 7605 7606 7607 7608 7609 7610 7611 7612 7613 7614 7615 7616 7617 7618 7619 7620 7621 7622 7623 7624 7625 7626 7627 7628 7629 7630 7631 7632 7633 7634 7635 7636 7637 7638 7639 7640 7641 7642 7643 7644 7645 7646 7647 7648 7649 7650 7651 7652 7653 7654 7655 7656 7657 7658 7659 7660 7661 7662 7663 7664 7665 7666 7667 7668 7669 7670 7671 7672 7673 7674 7675 7676 7677 7678 7679 7680 7681 7682 7683 7684 7685 7686 7687 7688 7689 7690 7691 7692 7693 7694 7695 7696 7697 7698 7699 7700 7701 7702 7703 7704 7705 7706 7707 7708 7709 7710 7711 7712 7713 7714 7715 7716 7717 7718 7719 7720 7721 7722 7723 7724 7725 7726 7727 7728 7729 7730 7731 7732 7733 7734 7735 7736 7737 7738 7739 7740 7741 7742 7743 7744 7745 7746 7747 7748 7749 7750 7751 7752 7753 7754 7755 7756 7757 7758 7759 7760 7761 7762 7763 7764 7765 7766 7767 7768 7769 7770 7771 7772 7773 7774 7775 7776 7777 7778 7779 7780 7781 7782 7783 7784 7785 7786 7787 7788 7789 7790 7791 7792 7793 7794 7795 7796 7797 7798 7799 7800 7801 7802 7803 7804 7805 7806 7807 7808 7809 7810 7811 7812 7813 7814 7815 7816 7817 7818 7819 7820 7821 7822 7823 7824 7825 7826 7827 7828 7829 7830 7831 7832 7833 7834 7835 7836 7837 7838 7839 7840 7841 7842 7843 7844 7845 7846 7847 7848 7849 7850 7851 7852 7853 7854 7855 7856 7857 7858 7859 7860 7861 7862 7863 7864 7865 7866 7867 7868 7869 7870 7871 7872 7873 7874 7875 7876 7877 7878 7879 7880 7881 7882 7883 7884 7885 7886 7887 7888 7889 7890 7891 7892 7893 7894 7895 7896 7897 7898 7899 7900 7901 7902 7903 7904 7905 7906 7907 7908 7909 7910 7911 7912 7913 7914 7915 7916 7917 7918 7919 7920 7921 7922 7923 7924 7925 7926 7927 7928 7929 7930 7931 7932 7933 7934 7935 7936 7937 7938 7939 7940 7941 7942 7943 7944 7945 7946 7947 7948 7949 7950 7951 7952 7953 7954 7955 7956 7957 7958 7959 7960 7961 7962 7963 7964 7965 7966 7967 7968 7969 7970 7971 7972 7973 7974 7975 7976 7977 7978 7979 7980 7981 7982 7983 7984 7985 7986 7987 7988 7989 7990 7991 7992 7993 7994 7995 7996 7997 7998 7999 8000

129 GCCC-----AGAGGCGCTGATACATGCGCTAGGNNNNACNCGTGCNNNNTCG 172

Db 741 oProProGlyLeuGlyIvarGlyIThrSerSergIyProProProLeuGlyAlaIalysGlyIse 76

02 177 ANACTNCTNCAGCACATGATGATNGCGTCNCTGGGATGNGCCTGNCACAGATCAACT 23

Db 761 rasnalProProProProAlaGlyArgGly-----77

237 ACTGCAGACCANCNAAGGACCTTNGAGACAGAATGTTATTANNAACCNACCTACCTTATN 299

Db 774 ---ArgAlaSerLeuGlyLeuGlyArgGlyArgGlyValSerValProThrAlaIaIaPr 79

QY 297 CCTTTGGCAANGANACTNTTCTT 321

DB 192 OLYS LYS INTRALALEUYS PROLEU 800

extensin-like protein 3 + alpha

C:Species: Medicago sativa (alfalfa)  
C:Date: 15-Jul-1995 #Sequence revision 03-Nov-1995 #text change 24-Nov-1995

C;Accession: S53504  
R;Abrahams, S.; Hayes, C.M.; Watson, J.M.

A/Title: Expression patterns of three genes in the stem of Lucerne (Medicago sativa L.)

A/Accession: S53504

A: molecule type: mRNA  
A: Residues: 1-328 <ABP>

A: Cross-references: GB:U36120; NID:g5355585; PIDN:AAB41815.1; PID:g5355586  
C: Superfamily: proline-rich protein 3

Alignment Scores:

Pred. No.:	1.42	Length:	228
Score:	74.00	Matches:	26

Percent Similarity:	55.00%	Conservative:	33
Best Local Similarity:	24.53%	Mismatches:	38

DB:	2	3
Gaps:	2	3

US-09-880-711-328 (1-350) x S53504 (1-228)

27 CCCAGTAGCCCTGAAGTGTGGCCCCCGTACCCTCTGTAC-----CCAGGGAGTTAGCG 80

121 ProAlaserProProPhneserProProAlaThrProProAlaThrProPro 14

```
01 ccgcctccctccctccgcctcaccaccccccttcactgcatgccatggagcnc 14
```

[illegible]

dbp 159 rThrProa]aprcalaprcalalvsleuIvsSerIvsa]aprotThrleuI]aprcovalle 17

186 CAGCACATGATGATNGCGTCNCCTGGGAATGNGCCTGNCACAGATCAACCTTACTGCAGAC 24

```

Db      179 userProSerAspaLaProLaProGlyLeuSerLeuSerProSerIleSerProSe 19
          |||  ::  ::  |||:::  ::  :::

```





**OY** 204 GTCCGTGGGATGAGGCCGCACAGATC-----AACCT 236  
       |||||:|||||:|||||:|||||:|||||:|||||  
**Db** 342 lThrValSerThrProLeuHisGlnValSerProThrGlyLeuAluProSerHisSerLe 362  
**OY** 237 ACTGACGACCNCANNAAGG-----ACCTTNGAGACAGAATGTATTANNAAACCNACTAC 290  
       |||||:|||||:|||||:|||||:|||||:|||||  
**Db** 352 uLeuSerThrGluAlaLysLeuValSerAlaIalagLygIyProLeuProProValSerTh 382  
**OY** 291 CTTAATNCCTTTGCAAAANGAANAACNTTCTTGTAANAATNTTG 333  
       |||||:|||:|||||:|||||:|||||:|||||  
**Db** 352 fLeuThrAlaLeuHisSerLeuGluGlnThrSerProGlyLeu 396

RESULT 15  
S39608

## N; Alterna

```

C:Date: 13-Jan1995 #sequence_revision 13-Jan-1995 #text_change 29-Aug-1997
C:Accession: S39608
R:Baeh, I.; Yaniv, M.
EMBO J. 12, 4229-4242, 1993
A:Title: More potent transcriptional activators or a transdominant inhibitor of the H
A:Reference number: S39607; MUID:94038905; PMID:7900959
A:Accession: S39608
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-542 <BAC>
A:Cross-references: EMBL:X71346
C:Genetics:
A:Gene: GDB:TCF1; HNF1; LEF1.
A:Cross-references: GDB:125297; OMM:142410; OMIM:600496
A:Map position: 12q24.3-12q24.3
C:Complex: homodimer; can also form heterodimers with, for example, HNF-1A form A
C:Function:
A:Description: transcription activator required for the expression of liver
A:Note: this form is a more potent transcriptional activator than the A form
C:Superfamily: transcription factor HNF-1; homeobox homology
C:Keywords: alternative splicing; DNA binding; homeobox; liver; nucleus; transcription
F:1-33/Region: dimerization
F:222-277/Domain: homeobox homology <HOX>

Alignment Scores:
Pred. No.:          3.51          Length:          542
Score:              70.00         Matches:          29
Percent Similarity: 53.91%       Conservative:     33
Best Local Similarity: 25.22%    Mismatches:      38
Query Match:        13.13%       Indels:          15
DB:                  2           Gaps:             3

```

QY 24 TACCCCG

[illegible]

QY 144 ATACATGCCCTAGGNNNNNCACNCCCTGACNNNTC

[illegible]

db 302 rleuthrAlaLeuHISerLeuGlnGlnThrSerProGlyLeu 396

Search completed: April 28, 2003, 15:38:39  
Job time : 23.5 secs

---



GenCore version 5.1.4.j5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: April 28, 2003, 15:33:39 ; Search time 12 seconds  
(without alignments)  
2419.452 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 533

Sequence: 1 atgagctacagctacgtca.....cttcagagattcaataaaaa 350

Scoring table:

BL0SUM62DX  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cn2.1/USPnc\_spool/US09880711/runat\_28042003\_112138\_843/app\_query.fasta.1.519  
-DB=SwissProt\_40 -OPMT=fastan -SUFPI=n2p.rsp -MINMATCH=0.1 -LOCPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62dx -TRANS=human40.cdt  
-LIST=45 -DOCCALGN=200 -THR\_SCORE=PCL -THR\_MAX=100 -THR\_MIN=0 -ALGN=15  
-MODE=LOCAL -OUTFMT=PEP -NGRM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=2000000000  
-USER=US09880711 @CGC 1.1.26. @runat.28042003\_112138\_843 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAR -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	240	45.0	197	1	RBMS_MOUSE
2	225.5	42.3	196	1	RBMS_MOUSE
3	151.5	28.4	196	1	RBMS_HUMAN
4	146.5	27.5	200	1	RBMS_XENLA
5	119	22.3	551	1	RBMS_CHICK
6	119	22.3	551	1	IFEA_HELPO
7	119	22.3	2477	1	IFEA_HELPO
8	117	22.0	1877	1	FINC_MOUSE
9	117	23.1	1877	1	PKCS_RAT
10	117	23.1	2477	1	PKCS_RAT
11	115	21.6	265	1	FINC_MOUSE
12	115	21.6	265	1	FINC_MOUSE
13	115	21.6	267	1	CR2A_LYCES
14	115	21.6	267	1	CR2E_LYCES
15	114	21.4	591	1	CR2F_LYCES
16	107	21.1	265	1	GAG_MMTVC
17	107	21.1	265	1	CR2A_LYCES
18	107	21.1	267	1	CR2C_LYCES
					PL4276 lycopersico

C 19	107	21.1	267	1	CR2F_LYCES	P14277 lycopersico
C 20	104.5	20.7	591	1	GAG_MMTVC	P11284 mouse mamma
C 21	92	17.3	371	1	CYB_ERYTA	O48085 eryx talari
C 22	91	16.0	371	1	CYB_ERYTA	O48085 eryx talari
C 23	78.5	14.7	2142	1	BAT2_HUMAN	P48634 homo sapien
C 24	73	13.7	631	1	INFA_HUMAN	P20823 homo sapien
C 25	71	14.3	371	1	CYB_ERYJA	O48076 eryx jaculu
C 26	70	14.8	371	1	CYB_ERYJA	O48076 eryx jaculu
C 27	68.5	12.9	2517	1	NSC2_HUMAN	O59518 h nuclear r
C 28	68	13.4	720	1	PSA4_MARMO	O59611 marsilea bo
C 29	66	14.0	233	1	SNQ4_MOUSE	O92112 mus musculu
C 30	65	12.2	356	1	HXB2_HUMAN	P14652 homo sapien
C 31	65	12.2	720	1	PSA4_MARMO	O59611 marsilea bo
C 32	65	12.2	1394	1	CNG4_BOVIN	O28181 bos laurus
C 33	64.5	12.1	338	1	OTX3_BRARE	O50267 brachydantio
C 34	64.5	12.1	751	1	FPL_MYTUA	O27409 mytilus gal
C 35	64	12.0	432	1	BRL4_EMENT	P10065 emeticeila
C 36	63.5	11.9	1163	1	RTM4_RAT	O91811 rattus norv
C 37	63	11.8	481	1	ETB2_HUMAN	O60883 homo sapien
C 38	63	11.8	512	1	GAG_SMSAV	P03330 simlan sarc
C 39	63	11.8	757	1	CIKE_HUMAN	O14003 homo sapien
C 40	63	11.8	1290	1	PER1_HUMAN	O15534 homo sapien
C 41	62.5	11.7	497	1	MAS2_HUMAN	O95645 homo sapien
C 42	62.5	11.7	591	1	MNT_MOUSE	O08789 mus musculu
C 43	62	11.6	87	1	CUO1_BACR	P80674 blaterus cr
C 44	62	11.6	520	1	RXR8_MOUSE	P28704 mus musculu
C 45	62	11.6	718	1	AT12_HSV11	P10230 herpes simp

#### ALIGNMENTS

RESULT 1  
RBMS\_MOUSE  
ID RBMS\_MOUSE STANDARD: PRT: 197 AA.

AC 09WB0;  
DT 16-OCT-2001 (rel. 40, Created)  
DT 16-OCT-2001 (rel. 40, Last annotation update)  
DE RNA-binding protein with multiple splicing (RBP-MS) (Heart, RRM  
DE Expressed Sequence) (Hermes).  
DE RBPMS OR HERMES.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart;  
RX MEDLINE=99195799; PubMed=1096065;  
RA Garber W.V., Yatskevych T.A., Antin P.B., Correia K.M., Conlon R.A.,  
RA Krieg P.A.;  
RT "The RNA-binding protein gene, hermes, is expressed at high levels in  
the developing heart.";  
RL Mech. Dev. 80:77-86(1999).  
RL  
CC -1- FUNCTION: MAY BIND RNA.  
CC -1- TISSUE SPECIFICITY: RNA EXPRESSED IN DEVELOPING HEART, WITH  
SIGNIFICANTLY HIGHER EXPRESSION IN THE ATRIA RELATIVE TO THE  
VENTRICLES.  
CC -1- DEVELOPMENTAL STAGE: RNA ALREADY DETECTED AT E9.5.  
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF148511; AAD39515.1; -.  
CC MD5: MG1:1334446; Rbpms.  
CC InterPro: IPR00504; RNA\_rec\_mot.  
CC Pfam: PF00076; rrm; 1.

[illegible]



**RESULT 6**

ID	FEA_HELPO	STANDARD:	PRT:	551 AA.
AC	P16274;			
DT	01-AUG-1990 (Rel. 15, Created)			
DR	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Non-neuronal cytoplasmic intermediate filament protein A (IFA) (fragments).			
OS	Helix pomatia (Roman snail) (Edible snail);			
OC	Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;			
OX	Helicoidea; Helicidae; Helix.			
RN	NCBL_TaxID=6536;			
RP	[1]			
RC	TISSUE=Oesophageal epithelium;			
RX	MEDLINE=89030591; PubMed=3181126;			
RA	Weber K., Plessmann U., Dodemont H., Kossmak-Strephan K.;			
RT	"Antibio acid sequences and homopolymer-forming ability of the intermediate filament proteins from an invertebrate epithellum.";			
RL	EMBO J. 7:2995-3001(1988).			
CC	- FUNCTION: Epithelial intermediate filament protein.			
CC	- SUBUNIT: Can form homopolymers.			
CC	- CELLULAR LOCATION: Cytoplasmic.			
CC	- ALTERNATIVE PRODUCTS: 2 isoforms, IFA (shown here) and IFB (AC p16275), are produced by alternative splicing.			
CC	- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.			
DR	PIR: S01295; S01295.			
DR	InterPro: IPRO01664; IF.			
DR	InterPro: IPRO01322; IF.tail.			
DR	Pfam: PF00038; Filament_4.			
DR	Pfam: PF00937; IF_tail; 1.			
KW	PROSITE; PS00226; IF; 1.			
KW	Intermediate filament; Coiled coil; Alternative splicing; Acetylation.			
FT	MOD_RES	1	HEAD.	
FT	DOMAIN	1..74	HEAD.	
FT	DOMAIN	75..423	ROD.	
FT	DOMAIN	424..551	TAIL.	
FT	DOMAIN	75..106	COIL_1A.	
FT	DOMAIN	107..120	LINKER_1.	
FT	DOMAIN	121..258	COIL_1B.	
FT	DOMAIN	259..275	LINKER_12.	
FT	DOMAIN	276..423	COIL_2.	
SQ	SEQUENCE	551 AA; 60943 MW; 116C48B6DAABAF9 CnC64;		

Alignment Scores:

Pred. No.:	Length:	Score:
1.86e-06	551	
119.00	Matches: 15	
46.79%	Conservative: 36	
Best local Similarity: 13.76%	Mismatches: 42	
Query Match: 23.52%	Indels: 16	
DB: 1	Gaps: 2	

```

US-09-880-711-328 (1-350) x FEA_HELPO (1-551)

OY   339 ATCTCAAAATNTTTCGAAGAACGTTNNCTTGCAAAA-----          298
      ||| |:::||||::::||:::||||:::||||:::||||:::||||:::
Db   447 lIeGlInSeRhThrAlaSerGIySerMeThValGlnatGSerSerLySIlyProlle 466
      ::::::::::::::::::::::::::::
OY   297 -----GNATAAGCTACTNGGTNTNTTAATCATCTGCTCNNAAGCT 256.
      :::: |||:::||||:::||||:::||||:::||||:::||||:::
Db   467 AlaphaeNsErValAsplnSeRISerValIlleGluaStHthrTSerCl y 486
      ::: |||:::||||:::||||:::||||:::||||:::||||:::
OY   255 CCCATTNCNTGGTGCAGTAGTGTTGATCTGTGNCAGGCCATCCCAAGACG NATCAT 196
      ::::: |||:::||||:::||||:::||||:::||||:::||||:::
Db   487 AlArgrAlaLystrHGInserLeuLIySGITyPArcValAsplyStrValAIagIyArg 506
      ::::: |||:::||||:::||||:::||||:::||||:::||||:::
OY   195 NCATGCTCGTNAGACGNTCCANNNTAGGNGCTGNNNNCTAGGACATGTFACAGGCC 136
      ::: |||:::||||:::||||:::||||:::||||:::||||:::
Db   507 ValAlalaSerlIleGlIUeuLySaSnTyfRaSplauPrOPaaStfHrLYstyTrtHlle 526
      ::::: |||:::||||:::||||:::||||:::||||:::||||:::
OY   135 TCtGGcGATCaCTGaAgGgGgtTaGtgTGAAGAAGCGGAGGAGGAGGAAGAcAGCcCTTA 76
      ::::: |||:::||||:::||||:::||||:::||||:::||||:::
    
```



Dh 527 TTPAlaIaIaGlyAlaIaIys-----AspArgAlaIthAlaIaspAsnIuGluIle 542  
Oy 75 ACTCCGCTCGCTACAGAGCTACGGG 49  
Db 543 AlaIaspMethSerLeuGlyAlaIcIy 551

RESULT 7  
FINC\_MOUSE STANDARD: PRT: 2477 AA.  
AC P11276: 061568: 061569: 061567: 064233;  
DT 01-JUL-1989 (rel. 11, Created)  
DT 01-NOV-1997 (rel. 35, Last sequence update)  
DT 15-JUN-2002 (rel. 41, Last annotation update)  
DE Fibronectin precursor (FN) (Fragments).  
GN FN1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN 11  
RP SEQUENCE OF 1-28 FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=94131313; PubMed=8299972;  
RA Polly P., Nicholson R.C.;  
RT "Sequence of the mouse fibronectin-encoding gene promoter region."  
RL Gene 137:353-354(1993).  
RN 12  
RP SEQUENCE OF 562-834 FROM N.A.  
RC STRAIN=SMR1;  
RX MEDLINE=95403556; PubMed=7673336;  
RA Tates J.F., Weller A., Timpi R., Ekblom M., Ekblom P.;  
RT "Regulation of mesenchymal extracellular matrix protein synthesis by  
RT transforming growth factor-beta and glucocorticoids in tumor  
RT stroma."  
RL J. Cell Sci. 108:2153-2162(1995).  
RN 13  
RP SEQUENCE OF 899-2376 FROM N.A.  
RA Gorski G., Aros M., Norton P.;  
RT Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.  
RN 14  
RP SEQUENCE OF 2375-2477 FROM N.A.  
RX MEDLINE=88124987; PubMed=3124113;  
RA Blatti S.P., Foster D.N., Ranganathan G., Moses H.L., Gelz M.J.;  
RT "Induction of fibronectin gene transcription and mRNA is a primary  
RT response to growth-factor stimulation of AKR-2B cells."  
RL Proc. Natl. Acad. Sci. U.S.A. 85:1119-1123(1988).  
RN 15  
RP SEQUENCE OF 2375-2477 FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=9311702; PubMed=1327855;  
RA Khadjian E.W., Salomon G., Leonard N., Tremblay S., Turlet H.;  
RT "Fibronectin gene expression in proliferating, quiescent, and SV40-  
RT infected mouse kidney cells."  
RL Exp. Cell Res. 202:464-470(1992).  
RN 16  
RP STRUCTURE BY NMR OF 1447-1630.  
RX MEDLINE=98202578; PubMed=9533887;  
RA Copie V., Tomita Y., Akiyama S.K., Aota S., Yamada K.M., Venable K.M.,  
RA Pastor R.W., Krueger S., Torchia D.A.;  
RT "Solution structure and dynamics of linked cell attachment modules of  
RT mouse fibronectin containing the RGD and synergy regions: comparison  
RT with the human fibronectin crystal structure."  
RL J. Mol. Biol. 277:663-682(1998).  
RN 17  
RP -1- FUNCTION: FIBRONECTIN BIND CELL SURFACES AND VARIOUS COMPOUNDS  
CC INCLUDING COLLAGEN, FIBRIN, HEPARIN, DNA, AND ACTIN. FIBRONECTINS  
CC ARE INVOLVED IN CELL ADHESION, CELL MOTILITY, GPMOIZATION, WOUND  
CC HEALING, AND MAINTENANCE OF CELL SHAPE.  
CC -1- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED  
CC VARIANTS, CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS.  
CC -1- ALTERNATIVE PRODUCTS: EACH OF THE "EXTRA DOMAIN" & THE CONNECTING  
CC STRAND 3 ARE PRESENT IN SOME FORMS OF FIBRONECTIN AND ABSENT IN  
CC OTHERS. THESE DIFFERENCES ARE DUE TO ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: PLASMA FN (SOLUBLE DIMERIC FORM) IS SECRETED  
CC BY HEPATOCYTES, CELLULAR FN (DIMERIC OR CROSS-LINKED MULTIMERIC  
CC FORMS), MADE BY FIBROBLASTS, EPITHELIAL AND OTHER CELL TYPES, IS  
CC DEPOSITED AS FIBRILS IN THE EXTRACELLULAR MATRIX.  
CC -1- PTM: SPLICED (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 12 FIBRONECTIN TYPE 1 DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE 11 DOMAINS.  
CC -1- SIMILARITY: CONTAINS 17 FIBRONECTIN TYPE 11 DOMAINS.  
CC -----  
CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z22729; CAA80422.1; -;  
DR EMBL: X82402; CAA57756.1; -;  
DR EMBL: X93167; CAA63654.1; -;  
DR EMBL: M18194; AAA37636.1; -;  
DR EMBL: S45680; AAB23491.1; -;  
DR PIR: A31371; A31371.  
DR PIR: G60597; G60597.  
DR PDB: 1HFN; 2F-APR-98.  
DR PDB: 2MFN; 2F-APR-98.  
DR MCD: M01:95566; P01.  
DR InterPro: IP0000561; EGF-like.  
DR InterPro: IP0003961; FN\_11.  
DR InterPro: IP0000562; FN\_type\_11.  
DR InterPro: IP0000083; Fibnctn.  
DR InterPro: IP0003962; FN11\_repeat.  
DR Pfam: PF00039; fn1; 4.  
DR Pfam: PF00041; fn3; 17.  
DR PRINTS: PR00014; FNTYPE111.  
DR SMART: SM00058; FN1; 4.  
DR SMART: SM00060; FN3; 14.  
DR PROSITE: PS00022; EGF\_1; 1.  
DR PROSITE: PS00023; FIBRONECTIN\_2; PARTIAL.  
DR PROSITE: PS01253; FIBRONECTIN\_1; 4.  
DR KEGG: Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;  
KW Sulfation; Cell adhesion; Repeat; Alternative splicing; Signal;  
KW 3D-structure.  
FT SIGNAL 1 32  
FT CHAIN 33 2477  
FT DOMAIN 53 273  
FT DOMAIN 308 608  
FT DNA\_BIND 906 1171  
FT DOMAIN 1357 1630  
FT DOMAIN 1811 2081  
FT DOMAIN 2265 2437  
FT DOMAIN 51 56  
FT DOMAIN 96 140  
FT DOMAIN 140 185  
FT DOMAIN 185 230  
FT DOMAIN 230 272  
FT DOMAIN 305 343  
FT DOMAIN 345 404  
FT DOMAIN 405 469  
FT DOMAIN 469 516  
FT DOMAIN 516 559  
FT DOMAIN 559 602  
FT DOMAIN 609 706  
FT DOMAIN 707 808  
FT DOMAIN 809 903  
FT DOMAIN 904 994  
FT DOMAIN 995 1084  
FT DOMAIN 1085 1172  
FT DOMAIN 1173 1264  
FT DOMAIN 1265 1355  
FT DOMAIN 1356 1447  
FT DOMAIN 1447 1536  
FT DOMAIN 1537 1630  
CC BY SIMILARITY.  
CC FIBRONECTIN.  
CC FIBRIN- AND HEPARIN-BINDING 1.  
CC COLLAGEN-BINDING.  
CC CELL-ATTACHMENT.  
CC HEPARIN-BINDING 2.  
CC FIBRIN-BINDING 2.  
CC FIBRONECTIN TYPE-1 1.  
CC FIBRONECTIN TYPE-1 2.  
CC FIBRONECTIN TYPE-1 3.  
CC FIBRONECTIN TYPE-1 4.  
CC FIBRONECTIN TYPE-1 5.  
CC FIBRONECTIN TYPE-1 6.  
CC FIBRONECTIN TYPE-1 1.  
CC FIBRONECTIN TYPE-11 1.  
CC FIBRONECTIN TYPE-11 2.  
CC FIBRONECTIN TYPE-11 3.  
CC FIBRONECTIN TYPE-11 4.  
CC FIBRONECTIN TYPE-11 5.  
CC FIBRONECTIN TYPE-11 6.  
CC FIBRONECTIN TYPE-11 7.  
CC FIBRONECTIN TYPE-11 8.  
CC FIBRONECTIN TYPE-11 9.  
CC FIBRONECTIN TYPE-11 10.  
CC FIBRONECTIN TYPE-11 11.  
CC FIBRONECTIN TYPE-11 12.  
CC FIBRONECTIN TYPE-11 13.  
CC FIBRONECTIN TYPE-11 14.  
CC FIBRONECTIN TYPE-11 15.  
CC FIBRONECTIN TYPE-11 16.  
CC FIBRONECTIN TYPE-11 17.  
CC FIBRONECTIN TYPE-11 18.  
CC FIBRONECTIN TYPE-11 19.  
CC FIBRONECTIN TYPE-11 20.  
CC FIBRONECTIN TYPE-11 21.  
CC FIBRONECTIN TYPE-11 22.  
CC FIBRONECTIN TYPE-11 23.  
CC FIBRONECTIN TYPE-11 24.  
CC FIBRONECTIN TYPE-11 25.  
CC FIBRONECTIN TYPE-11 26.  
CC FIBRONECTIN TYPE-11 27.  
CC FIBRONECTIN TYPE-11 28.  
CC FIBRONECTIN TYPE-11 29.  
CC FIBRONECTIN TYPE-11 30.  
CC FIBRONECTIN TYPE-11 31.  
CC FIBRONECTIN TYPE-11 32.  
CC FIBRONECTIN TYPE-11 33.  
CC FIBRONECTIN TYPE-11 34.  
CC FIBRONECTIN TYPE-11 35.  
CC FIBRONECTIN TYPE-11 36.  
CC FIBRONECTIN TYPE-11 37.  
CC FIBRONECTIN TYPE-11 38.  
CC FIBRONECTIN TYPE-11 39.  
CC FIBRONECTIN TYPE-11 40.  
CC FIBRONECTIN TYPE-11 41.  
CC FIBRONECTIN TYPE-11 42.  
CC FIBRONECTIN TYPE-11 43.  
CC FIBRONECTIN TYPE-11 44.  
CC FIBRONECTIN TYPE-11 45.  
CC FIBRONECTIN TYPE-11 46.  
CC FIBRONECTIN TYPE-11 47.  
CC FIBRONECTIN TYPE-11 48.  
CC FIBRONECTIN TYPE-11 49.  
CC FIBRONECTIN TYPE-11 50.  
CC FIBRONECTIN TYPE-11 51.  
CC FIBRONECTIN TYPE-11 52.  
CC FIBRONECTIN TYPE-11 53.  
CC FIBRONECTIN TYPE-11 54.  
CC FIBRONECTIN TYPE-11 55.  
CC FIBRONECTIN TYPE-11 56.  
CC FIBRONECTIN TYPE-11 57.  
CC FIBRONECTIN TYPE-11 58.  
CC FIBRONECTIN TYPE-11 59.  
CC FIBRONECTIN TYPE-11 60.  
CC FIBRONECTIN TYPE-11 61.  
CC FIBRONECTIN TYPE-11 62.  
CC FIBRONECTIN TYPE-11 63.  
CC FIBRONECTIN TYPE-11 64.  
CC FIBRONECTIN TYPE-11 65.  
CC FIBRONECTIN TYPE-11 66.  
CC FIBRONECTIN TYPE-11 67.  
CC FIBRONECTIN TYPE-11 68.  
CC FIBRONECTIN TYPE-11 69.  
CC FIBRONECTIN TYPE-11 70.  
CC FIBRONECTIN TYPE-11 71.  
CC FIBRONECTIN TYPE-11 72.  
CC FIBRONECTIN TYPE-11 73.  
CC FIBRONECTIN TYPE-11 74.  
CC FIBRONECTIN TYPE-11 75.  
CC FIBRONECTIN TYPE-11 76.  
CC FIBRONECTIN TYPE-11 77.  
CC FIBRONECTIN TYPE-11 78.  
CC FIBRONECTIN TYPE-11 79.  
CC FIBRONECTIN TYPE-11 80.  
CC FIBRONECTIN TYPE-11 81.  
CC FIBRONECTIN TYPE-11 82.  
CC FIBRONECTIN TYPE-11 83.  
CC FIBRONECTIN TYPE-11 84.  
CC FIBRONECTIN TYPE-11 85.  
CC FIBRONECTIN TYPE-11 86.  
CC FIBRONECTIN TYPE-11 87.  
CC FIBRONECTIN TYPE-11 88.  
CC FIBRONECTIN TYPE-11 89.  
CC FIBRONECTIN TYPE-11 90.  
CC FIBRONECTIN TYPE-11 91.  
CC FIBRONECTIN TYPE-11 92.  
CC FIBRONECTIN TYPE-11 93.  
CC FIBRONECTIN TYPE-11 94.  
CC FIBRONECTIN TYPE-11 95.  
CC FIBRONECTIN TYPE-11 96.  
CC FIBRONECTIN TYPE-11 97.  
CC FIBRONECTIN TYPE-11 98.  
CC FIBRONECTIN TYPE-11 99.  
CC FIBRONECTIN TYPE-11 100.  
CC FIBRONECTIN TYPE-11 101.  
CC FIBRONECTIN TYPE-11 102.  
CC FIBRONECTIN TYPE-11 103.  
CC FIBRONECTIN TYPE-11 104.  
CC FIBRONECTIN TYPE-11 105.  
CC FIBRONECTIN TYPE-11 106.  
CC FIBRONECTIN TYPE-11 107.  
CC FIBRONECTIN TYPE-11 108.  
CC FIBRONECTIN TYPE-11 109.  
CC FIBRONECTIN TYPE-11 110.  
CC FIBRONECTIN TYPE-11 111.  
CC FIBRONECTIN TYPE-11 112.  
CC FIBRONECTIN TYPE-11 113.  
CC FIBRONECTIN TYPE-11 114.  
CC FIBRONECTIN TYPE-11 115.  
CC FIBRONECTIN TYPE-11 116.  
CC FIBRONECTIN TYPE-11 117.  
CC FIBRONECTIN TYPE-11 118.  
CC FIBRONECTIN TYPE-11 119.  
CC FIBRONECTIN TYPE-11 120.  
CC FIBRONECTIN TYPE-11 121.  
CC FIBRONECTIN TYPE-11 122.  
CC FIBRONECTIN TYPE-11 123.  
CC FIBRONECTIN TYPE-11 124.  
CC FIBRONECTIN TYPE-11 125.  
CC FIBRONECTIN TYPE-11 126.  
CC FIBRONECTIN TYPE-11 127.  
CC FIBRONECTIN TYPE-11 128.  
CC FIBRONECTIN TYPE-11 129.  
CC FIBRONECTIN TYPE-11 130.  
CC FIBRONECTIN TYPE-11 131.  
CC FIBRONECTIN TYPE-11 132.  
CC FIBRONECTIN TYPE-11 133.  
CC FIBRONECTIN TYPE-11 134.  
CC FIBRONECTIN TYPE-11 135.  
CC FIBRONECTIN TYPE-11 136.  
CC FIBRONECTIN TYPE-11 137.  
CC FIBRONECTIN TYPE-11 138.  
CC FIBRONECTIN TYPE-11 139.  
CC FIBRONECTIN TYPE-11 140.  
CC FIBRONECTIN TYPE-11 141.  
CC FIBRONECTIN TYPE-11 142.  
CC FIBRONECTIN TYPE-11 143.  
CC FIBRONECTIN TYPE-11 144.  
CC FIBRONECTIN TYPE-11 145.  
CC FIBRONECTIN TYPE-11 146.  
CC FIBRONECTIN TYPE-11 147.  
CC FIBRONECTIN TYPE-11 148.  
CC FIBRONECTIN TYPE-11 149.  
CC FIBRONECTIN TYPE-11 150.  
CC FIBRONECTIN TYPE-11 151.  
CC FIBRONECTIN TYPE-11 152.  
CC FIBRONECTIN TYPE-11 153.  
CC FIBRONECTIN TYPE-11 154.  
CC FIBRONECTIN TYPE-11 155.  
CC FIBRONECTIN TYPE-11 156.  
CC FIBRONECTIN TYPE-11 157.  
CC FIBRONECTIN TYPE-11 158.  
CC FIBRONECTIN TYPE-11 159.  
CC FIBRONECTIN TYPE-11 160.  
CC FIBRONECTIN TYPE-11 161.  
CC FIBRONECTIN TYPE-11 162.  
CC FIBRONECTIN TYPE-11 163.  
CC FIBRONECTIN TYPE-11 164.  
CC FIBRONECTIN TYPE-11 165.  
CC FIBRONECTIN TYPE-11 166.  
CC FIBRONECTIN TYPE-11 167.  
CC FIBRONECTIN TYPE-11 168.  
CC FIBRONECTIN TYPE-11 169.  
CC FIBRONECTIN TYPE-11 170.  
CC FIBRONECTIN TYPE-11 171.  
CC FIBRONECTIN TYPE-11 172.  
CC FIBRONECTIN TYPE-11 173.  
CC FIBRONECTIN TYPE-11 174.  
CC FIBRONECTIN TYPE-11 175.  
CC FIBRONECTIN TYPE-11 176.  
CC FIBRONECTIN TYPE-11 177.  
CC FIBRONECTIN TYPE-11 178.  
CC FIBRONECTIN TYPE-11 179.  
CC FIBRONECTIN TYPE-11 180.  
CC FIBRONECTIN TYPE-11 181.  
CC FIBRONECTIN TYPE-11 182.  
CC FIBRONECTIN TYPE-11 183.  
CC FIBRONECTIN TYPE-11 184.  
CC FIBRONECTIN TYPE-11 185.  
CC FIBRONECTIN TYPE-11 186.  
CC FIBRONECTIN TYPE-11 187.  
CC FIBRONECTIN TYPE-11 188.  
CC FIBRONECTIN TYPE-11 189.  
CC FIBRONECTIN TYPE-11 190.  
CC FIBRONECTIN TYPE-11 191.  
CC FIBRONECTIN TYPE-11 192.  
CC FIBRONECTIN TYPE-11 193.  
CC FIBRONECTIN TYPE-11 194.  
CC FIBRONECTIN TYPE-11 195.  
CC FIBRONECTIN TYPE-11 196.  
CC FIBRONECTIN TYPE-11 197.  
CC FIBRONECTIN TYPE-11 198.  
CC FIBRONECTIN TYPE-11 199.  
CC FIBRONECTIN TYPE-11 200.  
CC FIBRONECTIN TYPE-11 201.  
CC FIBRONECTIN TYPE-11 202.  
CC FIBRONECTIN TYPE-11 203.  
CC FIBRONECTIN TYPE-11 204.  
CC FIBRONECTIN TYPE-11 205.  
CC FIBRONECTIN TYPE-11 206.  
CC FIBRONECTIN TYPE-11 207.  
CC FIBRONECTIN TYPE-11 208.  
CC FIBRONECTIN TYPE-11 209.  
CC FIBRONECTIN TYPE-11 210.  
CC FIBRONECTIN TYPE-11 211.  
CC FIBRONECTIN TYPE-11 212.  
CC FIBRONECTIN TYPE-11 213.  
CC FIBRONECTIN TYPE-11 214.  
CC FIBRONECTIN TYPE-11 215.  
CC FIBRONECTIN TYPE-11 216.  
CC FIBRONECTIN TYPE-11 217.  
CC FIBRONECTIN TYPE-11 218.  
CC FIBRONECTIN TYPE-11 219.  
CC FIBRONECTIN TYPE-11 220.  
CC FIBRONECTIN TYPE-11 221.  
CC FIBRONECTIN TYPE-11 222.  
CC FIBRONECTIN TYPE-11 223.  
CC FIBRONECTIN TYPE-11 224.  
CC FIBRONECTIN TYPE-11 225.  
CC FIBRONECTIN TYPE-11 226.  
CC FIBRONECTIN TYPE-11 227.  
CC FIBRONECTIN TYPE-11 228.  
CC FIBRONECTIN TYPE-11 229.  
CC FIBRONECTIN TYPE-11 230.  
CC FIBRONECTIN TYPE-11 231.  
CC FIBRONECTIN TYPE-11 232.  
CC FIBRONECTIN TYPE-11 233.  
CC FIBRONECTIN TYPE-11 234.  
CC FIBRONECTIN TYPE-11 235.  
CC FIBRONECTIN TYPE-11 236.  
CC FIBRONECTIN TYPE-11 237.  
CC FIBRONECTIN TYPE-11 238.  
CC FIBRONECTIN TYPE-11 239.  
CC FIBRONECTIN TYPE-11 240.  
CC FIBRONECTIN TYPE-11 241.  
CC FIBRONECTIN TYPE-11 242.  
CC FIBRONECTIN TYPE-11 243.  
CC FIBRONECTIN TYPE-11 244.  
CC FIBRONECTIN TYPE-11 245.  
CC FIBRONECTIN TYPE-11 246.  
CC FIBRONECTIN TYPE-11 247.  
CC FIBRONECTIN TYPE-11 248.  
CC FIBRONECTIN TYPE-11 249.  
CC FIBRONECTIN TYPE-11 250.  
CC FIBRONECTIN TYPE-11 251.  
CC FIBRONECTIN TYPE-11 252.  
CC FIBRONECTIN TYPE-11 253.  
CC FIBRONECTIN TYPE-11 254.  
CC FIBRONECTIN TYPE-11 255.  
CC FIBRONECTIN TYPE-11 256.  
CC FIBRONECTIN TYPE-11 257.  
CC FIBRONECTIN TYPE-11 258.  
CC FIBRONECTIN TYPE-11 259.  
CC FIBRONECTIN TYPE-11 260.  
CC FIBRONECTIN TYPE-11 261.  
CC FIBRONECTIN TYPE-11 262.  
CC FIBRONECTIN TYPE-11 263.  
CC FIBRONECTIN TYPE-11 264.  
CC FIBRONECTIN TYPE-11 265.  
CC FIBRONECTIN TYPE-11 266.  
CC FIBRONECTIN TYPE-11 267.  
CC FIBRONECTIN TYPE-11 268.  
CC FIBRONECTIN TYPE-11 269.  
CC FIBRONECTIN TYPE-11 270.  
CC FIBRONECTIN TYPE-11 271.  
CC FIBRONECTIN TYPE-11 272.  
CC FIBRONECTIN TYPE-11 273.  
CC FIBRONECTIN TYPE-11 274.  
CC FIBRONECTIN TYPE-11 275.  
CC FIBRONECTIN TYPE-11 276.  
CC FIBRONECTIN TYPE-11 277.  
CC FIBRONECTIN TYPE-11 278.  
CC FIBRONECTIN TYPE-11 279.  
CC FIBRONECTIN TYPE-11 280.  
CC FIBRONECTIN TYPE-11 281.  
CC FIBRONECTIN TYPE-11 282.  
CC FIBRONECTIN TYPE-11 283.  
CC FIBRONECTIN TYPE-11 284.  
CC FIBRONECTIN TYPE-11 285.  
CC FIBRONECTIN TYPE-11 286.  
CC FIBRONECTIN TYPE-11 287.  
CC FIBRONECTIN TYPE-11 288.  
CC FIBRONECTIN TYPE-11 289.  
CC FIBRONECTIN TYPE-11 290.  
CC FIBRONECTIN TYPE-11 291.  
CC FIBRONECTIN TYPE-11 292.  
CC FIBRONECTIN TYPE-11 293.  
CC FIBRONECTIN TYPE-11 294.  
CC FIBRONECTIN TYPE-11 295.  
CC FIBRONECTIN TYPE-11 296.  
CC FIBRONECTIN TYPE-11 297.  
CC FIBRONECTIN TYPE-11 298.  
CC FIBRONECTIN TYPE-11 299.  
CC FIBRONECTIN TYPE-11 300.  
CC FIBRONECTIN TYPE-11 301.  
CC FIBRONECTIN TYPE-11 302.  
CC FIBRONECTIN TYPE-11 303.  
CC FIBRONECTIN TYPE-11 304.  
CC FIBRONECTIN TYPE-11 305.  
CC FIBRONECTIN TYPE-11 306.  
CC FIBRONECTIN TYPE-11 307.  
CC FIBRONECTIN TYPE-11 308.  
CC FIBRONECTIN TYPE-11 309.  
CC FIBRONECTIN TYPE-11 310.  
CC FIBRONECTIN TYPE-11 311.  
CC FIBRONECTIN TYPE-11 312.  
CC FIBRONECTIN TYPE-11 313.  
CC FIBRONECTIN TYPE-11 314.  
CC FIBRONECTIN TYPE-11 315.  
CC FIBRONECTIN TYPE-11 316.  
CC FIBRONECTIN TYPE-11 317.  
CC FIBRONECTIN TYPE-11 318.  
CC FIBRONECTIN TYPE-11 319.  
CC FIBRONECTIN TYPE-11 320.  
CC FIBRONECTIN TYPE-11 321.  
CC FIBRONECTIN TYPE-11 322.  
CC FIBRONECTIN TYPE-11 323.  
CC FIBRONECTIN TYPE-11 324.  
CC FIBRONECTIN TYPE-11 325.  
CC FIBRONECTIN TYPE-11 326.  
CC FIBRONECTIN TYPE-11 327.  
CC FIBRONECTIN TYPE-11 328.  
CC FIBRONECTIN TYPE-11 329.  
CC FIBRONECTIN TYPE-11 330.  
CC FIBRONECTIN TYPE-11 331.  
CC FIBRONECTIN TYPE-11 332.  
CC FIBRONECTIN TYPE-11 333.  
CC FIBRONECTIN TYPE-11 334.  
CC FIBRONECTIN TYPE-11 335.  
CC FIBRONECTIN TYPE-11 336.  
CC FIBRONECTIN TYPE-11 337.  
CC FIBRONECTIN TYPE-11 338.  
CC FIBRONECTIN TYPE-11 339.  
CC FIBRONECTIN TYPE-11 340.  
CC FIBRONECTIN TYPE-11 341.  
CC FIBRONECTIN TYPE-11 342.  
CC FIBRONECTIN TYPE-11 343.  
CC FIBRONECTIN TYPE-11 344.  
CC FIBRONECTIN TYPE-11 345.  
CC FIBRONECTIN TYPE-11 346.  
CC FIBRONECTIN TYPE-11 347.  
CC FIBRONECTIN TYPE-11 348.  
CC FIBRONECTIN TYPE-11 349.  
CC FIBRONECTIN TYPE-11 350.  
CC FIBRONECTIN TYPE-11 351.  
CC FIBRONECTIN TYPE-11 352.  
CC FIBRONECTIN TYPE-11 353.  
CC FIBRONECTIN TYPE-11 354.  
CC FIBRONECTIN TYPE-11 355.  
CC FIBRONECTIN TYPE-11 356.  
CC FIBRONECTIN TYPE-11 357.  
CC FIBRONECTIN TYPE-11 358.  
CC FIBRONECTIN TYPE-11 359.  
CC FIBRONECTIN TYPE-11 360.  
CC FIBRONECTIN TYPE-11 361.  
CC FIBRONECTIN TYPE-11 362.  
CC FIBRONECTIN TYPE-11 363.  
CC FIBRONECTIN TYPE-11 364.  
CC FIBRONECTIN TYPE-11 365.  
CC FIBRONECTIN TYPE-11 366.  
CC FIBRONECTIN TYPE-11 367.  
CC FIBRONECTIN TYPE-11 368.  
CC FIBRONECTIN TYPE-11 369.  
CC FIBRONECTIN TYPE-11 370.  
CC FIBRONECTIN TYPE-11 371.  
CC FIBRONECTIN TYPE-11 372.  
CC FIBRONECTIN TYPE-11 373.  
CC FIBRONECTIN TYPE-11 374.  
CC FIBRONECTIN TYPE-11 375.  
CC FIBRONECTIN TYPE-11 376.  
CC FIBRONECTIN TYPE-11 377.  
CC FIBRONECTIN TYPE-11 378.  
CC FIBRONECTIN TYPE-11 379.  
CC FIBRONECTIN TYPE-11 380.  
CC FIBRONECTIN TYPE-11 381.  
CC FIBRONECTIN TYPE-11 382.  
CC FIBRONECTIN TYPE-11 383.  
CC FIBRONECTIN TYPE-11 384.  
CC FIBRONECTIN TYPE-11 385.  
CC FIBRONECTIN TYPE-11 386.  
CC FIBRONECTIN TYPE-11 387.  
CC FIBRONECTIN TYPE-11 388.  
CC FIBRONECTIN TYPE-11 389.  
CC FIBRONECTIN TYPE-11 390.  
CC FIBRONECTIN TYPE-11 391.  
CC FIBRONECTIN TYPE-11 392.  
CC FIBRONECTIN TYPE-11 393.  
CC FIBRONECTIN TYPE-11 394.  
CC FIBRONECTIN TYPE-11 395.  
CC FIBRONECTIN TYPE-11 396.  
CC FIBRONECTIN TYPE-11 397.  
CC FIBRONECTIN TYPE-11 398.  
CC FIBRONECTIN TYPE-11 399.  
CC FIBRONECTIN TYPE-11 400.  
CC FIBRONECTIN TYPE-11 401.  
CC FIBRONECTIN TYPE-11 402.  
CC FIBRONECTIN TYPE-11 403.  
CC FIBRONECTIN TYPE-11 404.  
CC FIBRONECTIN TYPE-11 405.  
CC FIBRONECTIN TYPE-11 406.  
CC FIBRONECTIN TYPE-11 407.  
CC FIBRONECTIN TYPE-11 408.  
CC FIBRONECTIN TYPE-11 409.  
CC FIBRONECTIN TYPE-11 410.  
CC FIBRONECTIN TYPE-11 411.  
CC FIBRONECTIN TYPE-11 412.  
CC FIBRONECTIN TYPE-11 413.  
CC FIBRONECTIN TYPE-11 414.  
CC FIBRONECTIN TYPE-11 415.  
CC FIBRONECTIN TYPE-11 416.  
CC FIBRONECTIN TYPE-11 417.  
CC FIBRONECTIN TYPE-11 418.  
CC FIBRONECTIN TYPE-11 419.  
CC FIBRONECTIN TYPE-11 420.  
CC FIBRONECTIN TYPE-11 421.  
CC FIBRONECTIN TYPE-11 422.  
CC FIBRONECTIN TYPE-11 423.  
CC FIBRONECTIN TYPE-11 424.  
CC FIBRONECTIN TYPE-11 425.  
CC FIBRONECTIN TYPE-11 426.  
CC FIBRONECTIN TYPE-11 427.  
CC FIBRONECTIN TYPE-11 428.  
CC FIBRONECTIN TYPE-11 429.  
CC FIBRONECTIN TYPE-11 430.  
CC FIBRONECTIN TYPE-11 431.  
CC FIBRONECTIN TYPE-11 432.  
CC FIBRONECTIN TYPE-11 433.  
CC FIBRONECTIN TYPE-11 434.  
CC FIBRONECTIN TYPE-11 435.  
CC FIBRONECTIN TYPE-11 436.  
CC FIBRONECTIN TYPE-11 437.  
CC FIBRONECTIN TYPE-11 438.  
CC FIBRONECTIN TYPE-11 439.  
CC FIBRONECTIN TYPE-11 440.  
CC FIBRONECTIN TYPE-11 441.  
CC FIBRONECTIN TYPE-11 442.  
CC FIBRONECTIN TYPE-11 443.  
CC FIBRONECTIN TYPE-11 444.  
CC FIBRONECTIN TYPE-11 445.  
CC FIBRONECTIN TYPE-11 446.  
CC FIBRONECTIN TYPE-11 447.  
CC FIBRONECTIN TYPE-11 448.  
CC FIBRONECTIN TYPE-11 449.  
CC FIBRONECTIN TYPE-11 450.  
CC FIBRONECTIN TYPE-11 451.  
CC FIBRONECTIN TYPE-11 452.  
CC FIBRONECTIN TYPE-11 453.  
CC FIBRONECTIN TYPE-11 454.  
CC FIBRONECTIN TYPE-11 455.  
CC FIBRONECTIN TYPE-11 456.  
CC FIBRONECTIN TYPE-11 457.  
CC FIBRONECTIN TYPE-11 458.  
CC FIBRONECTIN TYPE-11 459.  
CC FIBRONECTIN TYPE-11 460.  
CC FIBRONECTIN TYPE-11 461.  
CC FIBRONECTIN TYPE-11 462.  
CC FIBRONECTIN TYPE-11 463.  
CC FIBRONECTIN TYPE-11 464.  
CC FIBRONECTIN TYPE-11 465.  
CC FIBRONECTIN TYPE-11 466.  
CC FIBRONECTIN TYPE-11 467.  
CC FIBRONECTIN TYPE-11 468.  
CC FIBRONECTIN TYPE-11 469.  
CC FIBRONECTIN TYPE-11 470.  
CC FIBRONECTIN TYPE-11 471.  
CC FIBRONECTIN TYPE-11 472.  
CC FIBRONECTIN TYPE-11 473.  
CC FIBRONECTIN TYPE-11 474.  
CC FIBRONECTIN TYPE-11 475.  
CC FIBRONECTIN TYPE-11 476.  
CC FIBRONECTIN TYPE-11 477.  
CC FIBRONECTIN TYPE-11 478.  
CC FIBRONECTIN TYPE-11 479.  
CC FIBRONECTIN TYPE-11 480.  
CC FIBRONECTIN TYPE-11 481.  
CC FIBRONECTIN TYPE-11 482.  
CC FIBRONECTIN TYPE-11 483.  
CC FIBRONECTIN TYPE-11 484.  
CC FIBRONECTIN TYPE-11 485.  
CC FIBRONECTIN TYPE-11 486.  
CC FIBRONECTIN TYPE-11 487.  
CC FIBRONECTIN TYPE-11 488.  
CC FIBRONECTIN TYPE-11 489.  
CC FIBRONECTIN TYPE-11 490.  
CC FIBRONECTIN TYPE-11 491.  
CC FIBRONECTIN TYPE-11 492.  
CC FIBRONECTIN TYPE-11 493.  
CC FIBRONECTIN TYPE-11 494.  
CC FIBRONECTIN TYPE-11 495.  
CC FIBRONECTIN TYPE-11 496.  
CC FIBRONECTIN TYPE-11 497.  
CC FIBRONECTIN TYPE-11 498.  
CC FIBRONECTIN TYPE-11 499.  
CC FIBRONECTIN TYPE-11 500.  
CC FIBRONECTIN TYPE-11 501.  
CC FIBRONECTIN TYPE-11 502.  
CC FIBRONECTIN TYPE-11 503.  
CC FIBRONECTIN TYPE-11 504.  
CC FIBRONECTIN TYPE-11 505.  
CC FIBRONECTIN TYPE-11 506.  
CC FIBRONECTIN TYPE-11 507.  
CC FIBRONECTIN TYPE-11 508.  
CC FIBRONECTIN TYPE-11 509.  
CC FIBRONECTIN TYPE-11 510.  
CC FIBRONECTIN TYPE-11 511.  
CC FIBRONECTIN TYPE-11 512.  
CC FIBRONECTIN TYPE-11 513.  
CC FIBRONECTIN TYPE-11 514.  
CC FIBRONECTIN TYPE-11 515.  
CC FIBRONECTIN TYPE-11 516.  
CC FIBRONECTIN TYPE-11 517.  
CC FIBRONECTIN TYPE-11 518.  
CC FIBRONECTIN TYPE-11 519.  
CC FIBRONECTIN TYPE-11 520.  
CC FIBRONECTIN TYPE-11 521.  
CC FIBRONECTIN TYPE-11 522.  
CC FIBRONECTIN TYPE-11 523.  
CC FIBRONECTIN TYPE-11 524.  
CC FIBRONECTIN TYPE-11 525.

```

FT DOMAIN 1631 1720 FIBRONECTIN TYPE-III 12.
FT DOMAIN 1721 1810 FIBRONECTIN TYPE-III 13 (EXTRA DOMAIN 2).
FT DOMAIN 1811 1902 FIBRONECTIN TYPE-III 14.
FT DOMAIN 1903 1991 FIBRONECTIN TYPE-III 15.
FT DOMAIN 1992 2081 FIBRONECTIN TYPE-III 16.
FT DOMAIN 2082 2201 FIBRONECTIN STRAND 3 (CS-3) (V REGION).
FT DOMAIN 2202 2283 FIBRONECTIN TYPE-III 17.
FT DOMAIN 2284 2338 FIBRONECTIN TYPE-I 10.
FT DOMAIN 2339 2381 FIBRONECTIN TYPE-I 11.
FT DOMAIN 2383 2426 FIBRONECTIN TYPE-I 12.
FT SITE 1614 1616 CELL ATTACHMENT SITE.
FT SITE 2181 2183 CELL ATTACHMENT SITE.
FT DISULFID 561 589 BY SIMILARITY.
FT DISULFID 587 599 BY SIMILARITY.
FT DISULFID 2296 2325 BY SIMILARITY.
FT DISULFID 2323 2335 BY SIMILARITY.
FT DISULFID 2341 2368 BY SIMILARITY.
FT DISULFID 2366 2378 BY SIMILARITY.
FT DISULFID 2385 2409 BY SIMILARITY.
FT DISULFID 2407 2423 BY SIMILARITY.
FT DISULFID 2458 2458 INTERCHAIN (WITH 2462 OF OTHER CHAIN).
FT DISULFID 2462 2462 INTERCHAIN (WITH 2458 OF OTHER CHAIN).
FT CARBOHYD 1006 1006 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1243 1243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1290 1290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2198 2198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MOD_RES 2475 2475 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 2440 2440 N -> T (IN REF. 5).
SQ SEQUENCE 2477 AA; 271416 MW; 8D2B63B474E2A2M4 CRC64;

```

```

Alignment Scores:
Pred. No.: 1 75e-06
Score: 119.00 Length: 2477
Percent Similarity: 60.248 Matches: 18
Best Local Similarity: 21.698 Conservative: 32
Query Match: 22.338 Mismatches: 17
DB: 1 Indels: 17
Gaps: 4

```

US-09-880-711-328 (1-350) x FINC\_MOUSE (1-2477)

```

QY 100 CCGCTTACACCTACCCGCTTACGATGCCAGAGGCGCTGATACATGCTT-----153
    |||  :: |||||::| |||||  |||||  |||||  |||||  |||||  |||||
DB 2153 ProThraAlaIatHrProValArgLeuArgProArg-----ProTyrLeuProVal 2170
QY 154 -----AGNNNNNACCCGCGACNNNTGACNCTNCTGAC-CACATGATGAT 200
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2171 AsGclugluValInIleGlyHisValProArgGlyAspValAspTyrHisLeuTyrPro 2190
QY 201 NCGCTGCTGGGAATGCGCTGNCACAGATCAACCTACTGACGACGACGAGGAGCTT 260
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2191 HisValProGlyLeuAsnProAsn-----AlaSerThrGlyGlnGluAla 2205
QY 261 NMGACAGAGATGTATTANNACACCTACTATNCCCTTTGCAANAGMANACTTTCT 620
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2206 LeuSerGln-----ThrThrIleSerTyrThrProPheGlnIleuSerGlnTyr 2222
QY 321 TGGANATN 329
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2223 IleIleSer 2225

```

```

RESULT 8
PKCS_RAT STANDARD; PRT: 1877 AA.
AC P41413; 062914;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE (Proprotein convertase PC5) (Subtilisin/kexin-like protease PC5)
DE (Convertase PC5) (PC5) (PC6) (Fragments).
GN PCSK5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A. (SHORT ISOFORM), AND TISSUE-SPECIFICITY.
RX TISSUE-Adrenal gland;
RX MEDLINE=93342056; PubMed=8341687;
RA Lussan J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a
RT candidate proprotein convertase expressed in endocrine and
RT nonendocrine cells."
RT Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
RN [2]
SEQUENCE OF 1676-1877 FROM N.A. (LONG ISOFORM).
RP TISSUE-Adrenal gland.
RC De Bie I., Marcinkiewicz M., Nakayama K., Lazure C., Seidah N.G.;
RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
RN [3]
SUBCELLULAR LOCATION.
RX MEDLINE=20214819; PubMed=10749928;
RA Xiang Y., Molloy S.S., Thomas L., Thomas G.;
RT "The PC5B cytoplasmic domain contains two acidic clusters that direct
RT sorting to distinct trans-Golgi network/endosomal compartments."
RL Mol. Biol. Cell 11:1257-1273(2000).
RN [4]
DEVELOPMENTAL EXPRESSION.
RX MEDLINE=97166043; PubMed=9013936;
RA Zheng M., Seidah N.G., Pintar J.E.;
RT "The developmental expression in the rat CNS and peripheral tissues of
RT proteases PC5 and PC64 mRNAs: comparison with other proprotein
RT processing enzymes."
RL Dev. Biol. 181:268-283(1997).
CC -1- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
CC OF CLEAVAGE AT THE RX(R)/R CONSENSUS MOTIF. MAY BE RESPONSIBLE
CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED
CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION
CC OF GROWTH FACTORS.
CC -1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
CC PROPROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-ZAA BONDS, WHERE XAA
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR IYS.
CC -1- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED
CC SECRETORY PATHWAY. PC5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
CC EARLY ENDOSOMES (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PC5B/LONG (SHOWN HERE)
CC AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE INTESTINE, BRAIN, ADRENAL
CC GLAND, ANTERIOR PITUITARY, THYROID, OVARIES, TESTIS AND LUNG.
CC HIGHEST LEVELS ARE FOUND IN THE GUT, DUODENUM, JEJUNUM AND ILEUM.
CC EXPRESSION IS HIGHER IN FEMALE THAN IN MALE REPRODUCTIVE ORGANS.
CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED AT E9 IN HIGHLY RESTRICTED
CC REGIONS OF THE NEURAL TUBE, IN CAUDAL MYOTOMES, AND AT THE
CC MATERNO-EMBRYONIC JUNCTION OF THE UTERUS. AT E10, RESTRICTED
CC EXPRESSION IS DETECTED IN THE OPTIC AND OTIC VESICLES, THE ROOF OF
CC MIDBRAIN, AND TRUNK MYOTOMES. BY MIDGESTATION (E13-E16),
CC EXPRESSION IN THE DEVELOPING NERVOUS SYSTEM HAS EXPANDED TO
CC MULTIPLE REGIONS INCLUDING HIPPOCAMPUS, THALAMUS, HYPOTHALAMUS,
CC BRAIN STEM, AND SPINAL CORD. EXPRESSION IS ALSO DETECTED IN
CC SEVERAL PERIPHERAL ORGAN SYSTEMS, INCLUDING GUT, LUNG, ADRENAL AND
CC KIDNEY PRIMORDIA.
CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
CC ASSISTING THE FOLDING OF THE ZIMOGEN WITHIN THE ENDOPLASMIC
CC RETICULUM.
CC -1- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN
CC SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS
CC WITH THE TGN SORTING PROTEIN PACS-1.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



```

FT CARBOHYD      383   383   ) (POTENTIAL)
FT CARBOHYD      667   667   N-LINKED (GLCNAC . .) (POTENTIAL)
FT CARBOHYD      754   754   N-LINKED (GLCNAC . .) (POTENTIAL)
FT CARBOHYD      804   804   N-LINKED (GLCNAC . .) (POTENTIAL)
FT CARBOHYD      854   854   N-LINKED (GLCNAC . .) (POTENTIAL)
FT CARBOHYD     1710   1710   N-LINKED (GLCNAC . .) (POTENTIAL)
FT CARBOHYD     1732   1732   N-LINKED (GLCNAC . .) (POTENTIAL)
FT VASAPPLIC      878   915   XXXXXXXXXXXXXXXXXXNNCQKRVLDQLCCIKCTTPOG
                                *IN ISOFORM PCSA*)
FT FT             916   1877   MISSING (IN ISOFORM PCSA).
SQ SEQUENCE      1877 AA; 207888 MW; 890955DDC6053444A CRC64;

Alignment Scores:
Pred. No.:           3,06e-06          Length:         1877
Score:              117.00            Matches:        11
Percent Similarity: 63.41%            Conservative:   15
Best Local Similarity: 26.83%          Mismatches:    12
Query Match:       23.12%             Indels:        3
Db:                1.                  Gaps:         1

US-09-880-711-328 (1-350) x PKCS_RAT (1-1877)

OY 235 GGTCATCTGTGCAGCGNCATTCGCCAGNCGCATCATCATCATGGTGAGNAAGNTIC 176
DB 761 GlyLeuSerLeuGlnGly-----SerArgcysSeraValThrGySGIASpGlyGln 777
OY 175 CANNNGTCAGGNGGTGNMNNNCCTTAGGCATGTATACAGNCCCCTCGGCGCACTGAAGCG 116
DB 778 PhePheSerGIrghISAsPCyGSgInProCyShIsArgrPhcySaIatIrhcYSalIGLyAla 798
OY 115 GGG 113
DB 798 gHy 798

RESULT 10
FINC_MOUSE STANDARD; PRT: 2477 AA.
ID FINC_MOUSE AC PI1276; Q61568; Q61569; Q61567; Q64233;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fibronectin precursor (FN) (fragments).
GN FN1.
OC Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-28 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94131313; PubMed=8299972;
RA Polly P., Nicholson R.C.;
RT "Sequence of the mouse fibronectin-encoding gene promoter region.";
RL Gene 137:353-354(1993).
RN [2]
RP SEQUENCE OF 562-834 FROM N.A.
RC STRAIN=NMRI;
RX MEDLINE=95403556; PubMed=7673336;
RA Talts J.F., Weller A., Timpl R., Ekblom M., Ekblom P.;
RT "Regulation of mesenchymal extracellular matrix protein synthesis by
transferring growth factor-beta and glucocorticoids in tumor
stroma.";
RL J. Cell Sci. 108:2153-2162(1995).
RN [3]
RP SEQUENCE OF 899-2376 FROM N.A.
RC Gorski G., Aros W., Norton P.;
RA Submitted (DEC-1995) to the EMBL/GenBank/DDJB databases.
RN [4]
RP SEQUENCE OF 2375-2477 FROM N.A.
RX MEDLINE=88124987; PubMed=3124113;
RA Blatti S.P., Foster D.N., Rangathan G., Moses H.L., Getz M.J.;
RT "Induction of fibronectin gene transcription and mRNA is a primary

```

RT response to growth-factor stimulation of AKR-2B cells";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:1119-1123(1988).  
 RN (5)  
 RP SEQUENCE OF 2375-2477 FROM N.A.  
 RC TISSUE=Kidney;  
 KX MEDLINE=93011702; Pubmed=1327855;  
 KA Khandjian E.W., Salomon G., Leonard N., Tremblay S., Turter H.;  
 RT "Fibronectin gene expression in proliferating, quiescent, and SV40-  
 RT infected mouse kidney cells";  
 RL Exp. Cell Res. 202:464-470(1992).  
 RN (6)  
 RP STRUCTURE BY NMR OF 1447-1630.  
 RX MEDLINE=98202578; Pubmed=9533887;  
 RA Copie V., Tomita Y., Akiyama S.K., Aota S., Yamada K.M., Venable R.M.,  
 RA Pastor R.W., Krueger S., Torchia D.A.;  
 RT "solution structure and dynamics of linked cell attachment modules of  
 RT mouse fibronectin containing the RGD and synergy regions: comparison  
 RT with the human fibronectin crystal structure.";  
 RL J. Mol. Biol. 277:663-682(1998).  
 CC -1- FUNCTION: FIBRONECTIN BIND CELL SURFACES AND VARIOUS COMPOUNDS  
 CC INCLUDING COLLAGEN, FIBRIN, HEPARIN, DNA, AND ACTIN. FIBRONECTINS  
 CC ARE INVOLVED IN CELL ADHESION, CELL MOTILITY, OPSONIZATION, WOUND  
 CC HEALING, AND MAINTENANCE OF CELL SHAPE.  
 CC -1- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED  
 CC VARIANTS, CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS;  
 CC TO A LESSER EXTENT HOMODIMERS.  
 CC -1- ALTERNATIVE PRODUCTS: EACH OF THE "EXTRA DOMAIN" & "THE CONNECTING  
 CC STRAND 3 ARE PRESENT IN SOME FORMS OF FIBRONECTIN AND ABSENT IN  
 CC OTHERS. THESE DIFFERENCES ARE DUE TO ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: PLASMA FN (SOLUBLE DIMERIC FORM) IS SECRETED  
 CC BY HEPATOCYTES. CELLULAR FN (DIMERIC OR CROSS-LINKED MULTIMERIC  
 CC FORMS), MADE BY FIBROBLASTS, EPITHELIAL AND OTHER CELL TYPES, IS  
 CC DEPOSITED AS FIBRILS IN THE EXTRACELLULAR MATRIX.  
 CC -1- PTM: SULFATED (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE II DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 17 FIBRONECTIN TYPE III DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announc/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z23729; CAAB0422.1; -;  
 DR EMBL: X82402; CAAB7796.1; -;  
 DR EMBL: X93167; CAAB3634.1; -;  
 DR EMBL: M18194; AAA37636.1; -;  
 DR EMBL: S45680; AAB23491.1; -;  
 DR PIR: A31371; A31371;  
 DR PIR: C60597; C60597;  
 DR PDB: 1MPN; 29-APR-98;  
 DR PDB: 2MPN; 29-APR-98;  
 DR PDB: MGI; 95566; F01;  
 DR InterPro: IFR000561; EGF-like.  
 DR InterPro: IFR003561; FN\_III.  
 DR InterPro: IFR000562; FN\_Type\_II.  
 DR InterPro: IFR000083; fibnctn1.  
 DR InterPro: IFR003962; Phl11-repeat.  
 DR Pfam: PF00049; fn1; 4.  
 DR Pfam: PF00041; fn3; 17.  
 DR PRINTS: PROCO14; FNTYPE11.  
 DR SMART: SM00C58; FN1; 4.  
 DR SMART: SM00C60; FN3; 14.  
 DR PROSITE: PSC0022; EGF\_1; 1.  
 DR PROSITE: PSC0023; FIBRONECTIN\_2; PARTIAL.  
 DR PROSITE: PSC1253; FIBRONECTIN\_1; 4.  
 KM Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;  
 KM Sulfation; Cell adhesion; Repeat; Alternative splicing; Signal;  
 KM 3D-structure; 1  
 FT SIGNAL 1 32 BY SIMILARITY.

FT	CHAIN	33	2477	FIBRONECTIN.
FT	DOMAIN	53	273	FIBRIN- AND HEPARIN-BINDING 1.
FT	DOMAIN	308	608	COLLAGEN-BINDING.
FT	DNA_BIND	506	1171	CELL-ATTACHMENT.
FT	DOMAIN	1357	1630	HEPARIN-BINDING 2.
FT	DOMAIN	1811	2081	FIBRIN-BINDING 2.
FT	DOMAIN	2256	2427	FIBRONECTIN TYPE-1 1.
FT	DOMAIN	51	96	FIBRONECTIN TYPE-1 2.
FT	DOMAIN	56	140	FIBRONECTIN TYPE-1 3.
FT	DOMAIN	140	185	FIBRONECTIN TYPE-1 4.
FT	DOMAIN	185	230	FIBRONECTIN TYPE-1 5.
FT	DOMAIN	230	272	FIBRONECTIN TYPE-1 6.
FT	DOMAIN	306	343	FIBRONECTIN TYPE-1 7.
FT	DOMAIN	345	404	FIBRONECTIN TYPE-1 8.
FT	DOMAIN	405	469	FIBRONECTIN TYPE-1 9.
FT	DOMAIN	468	516	FIBRONECTIN TYPE-1 10.
FT	DOMAIN	516	559	FIBRONECTIN TYPE-1 11.
FT	DOMAIN	559	602	FIBRONECTIN TYPE-1 12.
FT	DOMAIN	609	706	FIBRONECTIN TYPE-1 13.
FT	DOMAIN	707	808	FIBRONECTIN TYPE-1 14.
FT	DOMAIN	809	903	FIBRONECTIN TYPE-1 15.
FT	DOMAIN	904	994	FIBRONECTIN TYPE-1 16.
FT	DOMAIN	995	1084	FIBRONECTIN TYPE-1 17.
FT	DOMAIN	1085	1172	FIBRONECTIN TYPE-1 18.
FT	DOMAIN	1173	1264	FIBRONECTIN TYPE-1 19.
FT	DOMAIN	1265	1355	FIBRONECTIN TYPE-1 20.
FT	DOMAIN	1356	1446	FIBRONECTIN TYPE-1 21.
FT	DOMAIN	1447	1536	FIBRONECTIN TYPE-1 22.
FT	DOMAIN	1537	1630	FIBRONECTIN TYPE-1 23.
FT	DOMAIN	1631	1720	FIBRONECTIN TYPE-1 24.
FT	DOMAIN	1721	1810	FIBRONECTIN TYPE-1 25.
FT	DOMAIN	1811	1902	FIBRONECTIN TYPE-1 26.
FT	DOMAIN	1903	1991	FIBRONECTIN TYPE-1 27.
FT	DOMAIN	1992	2081	FIBRONECTIN TYPE-1 28.
FT	DOMAIN	2082	2201	CONNECTING STAND 3 (CS-3) (V REGION).
FT	DOMAIN	2202	2283	FIBRONECTIN TYPE-1 10.
FT	DOMAIN	2294	2338	FIBRONECTIN TYPE-1 11.
FT	DOMAIN	2339	2381	FIBRONECTIN TYPE-1 12.
FT	DOMAIN	2383	2426	CELL ATTACHMENT SITE.
FT	SITE	1614	1616	CELL ATTACHMENT SITE.
FT	SITE	2181	2183	CELL ATTACHMENT SITE.
FT	DISULFID	561	589	BY SIMILARITY.
FT	DISULFID	587	599	BY SIMILARITY.
FT	DISULFID	2296	2325	BY SIMILARITY.
FT	DISULFID	2323	2335	BY SIMILARITY.
FT	DISULFID	2341	2368	BY SIMILARITY.
FT	DISULFID	2366	2378	BY SIMILARITY.
FT	DISULFID	2385	2409	BY SIMILARITY.
FT	DISULFID	2407	2423	BY SIMILARITY.
FT	DISULFID	2458	2458	INTERCHAIN (WITH 2458 OF OTHER CHAIN).
FT	DISULFID	2462	2462	INTERCHAIN (WITH 2462 OF OTHER CHAIN).
FT	CARBOHYD	1006	1006	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1243	1243	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1290	1290	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2198	2198	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	MOD_RES	2475	2475	PHOSPHORYLATION (BY SIMILARITY).
FT	CONFLICT	2440	2440	N->T (IN REF. 5).
SO	SEQUENCE	2477 AA;	271416 MW;	8D2H5B474E2A2M4 CIR664;

## Alignment Scores:

Pred. No.:	3.03e-06	Length:	2477
Score:	117.00	Matches:	20
Percent Similarity:	48.11%	Conservative:	31
Best Local Similarity:	18.87%	Mismatches:	35
Query Match:	23.12%	Indels:	20
DB:	1	Gaps:	3

US-09-880-711-328 (1-350) x F1NC\_MOUSE (1-2477)

OY 342 GAATTCCTCTCAAAANNTTCAGAAAGGTTTTCGAAAGAGGNTAGGTAGTTCG 283  
 DB 1403 GluLeuSer11LeuProSerAspAsnAlaVal1Val1euThrAsnLeuProGly1Thr 1422







CC -1- SUBUNIT: THE LHC COMPLEX CONSISTS OF CHLOROPHYLLS (A & B) AND  
CC CHLOROPIHLL A-B BINDING PROTEINS.  
CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.

CC This SWISS-PROT entry is copyright. It is produced through a collabo  
CC between the Swiss Institute of Bioinformatics and the EMBL Outsta  
CC the European Bioinformatics Institute. There are no restrictions o  
CC use by non-profit institutions as long as its content is in pub  
CC modified and this statement is not removed. Usage by and for comm  
CC entities requires a license agreement (See <http://www.isb-sib.ch/auth>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M30621; AAA34156.1; -  
DR EMBL; M30622; AAA34157.1; -  
DR InterPro; IPR001344; Chloro\_Abbind.  
DR Pfam; PF00504; chloroa\_b-bind.1.  
DR ProDom; PD000275; Chloro\_Abbind.1.  
KW Chlorophyll; Photosynthesis; Photosystem I; Photosystem II;  
KW Thylakoid; Membrane; Chloroplast; Transl peptide; Multigene family;  
FT TRANSMEMBRANE; Phosphorylation.  
FT TRANSIT 1 34 CHLOROPLAST.  
FT CHAIN 35 267 POTENTIAL.  
FT TRANSEM 153 173 POTENTIAL.  
FT TRANSEM 221 237 POTENTIAL.  
SO SEQUENCE 267 AA; 28440 MW; CEA88B8ACD873B65 CRC64;

Alignment Scores:  
Pred. NO.: 5.71e-06 Length: 267  
Score: 115.00 Matches: 9  
Percent Similarity: 56.52% Conservative: 17  
Best Local Similarity: 15.22% Mismatches: 20  
Query Match: 21.58% Indels: 0  
Db: 1 Gaps: 0

US-09-880-711-328 (1-350) x CBFE\_LYCES (1-267)

OY 130 CCCAGAGGCNCCTGATCATGATCGCTGAGNNNACACMCCTGCAGCNNTGACACTNCACGC 18  
||| |||::: ||| :::::::::::::::::::::  
DB 198 ProlenglyleAlaIagluAspProclualaphelaIaIguleuysVallysglluleys 21  
OY 190 ACATGATGATNGCGTCNCTGGGAATGNGCCTGCACAGATCAACTTACTGCAGACAC 24  
::: |||::: |||::: :::::::::::::::::::::  
DB 218 AsngcyArglueAlamelPheSermetPhecglyPhepevalGlalalaIvalThrly 23  
OY 250 NAAGSGAGCTTNNGAGAC 267  
DB 238 LysgLyProLeugluASN 243

RESULT 15  
GAG\_MMTVC STANDARD: PRT; 591 AA.  
ID GAG\_MMTVC  
AC P11284;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE GAG polypeptide [Contains: protein p10; Phosphorylated protein pP2];  
DE Protein p3; Protein p8; Major core protein p27; Nucleic acid binding  
DE protein p14].  
DE GAG.  
GN GAG.  
OS Mouse mammary tumor virus (strain C3H).  
OC Viruses; Retroid viruses; Retroviridae; Betaretrovirus.  
OX NCBI\_TaxId=11759;  
RN [1]  
RP SEQUENCE OF 359-591 FROM N.A.  
RX MEDLINE=87231993; PubMed=3035577;  
RA Jacks T., Townsend K., Varnus H.E., Majors J.;  
RT "Two efficient ribosomal frameshifting events are required for  
RL synthesis of mouse mammary tumor virus gag-related polypeptides.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:4298-4302(1987).  
RN [2]  
RP PARTIAL SEQUENCE. AND DOMAINS.  
RX MEDLINE=89259031; PubMed=2542570;  
RA Hizi A., Henderson L.E., Copeland T.D., Sowder R.C., Kruttsch H.C.,

```

RA      Oroszian S.;
RL      "Analysis of gag proteins from mouse mammary tumor virus.";
RL      J. Virol. 63:2543-2549(1989).
CC      -I- FUNCTION: P14 BINDS TO SINGLE STRANDED DNA.
CC      -I- SIMILARITY: CONTAINS 2 CCHC-TYPE ZINC FINGERS.
CC      -----
CC      This SWISS-PROT entry is copyrighted. It is produced through a collabor
CC      between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commer
CC      entities requires a license agreement (See http://www.isb-sib.ch/anno
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M16766; AAA66623.1;
DR      PIR; A29029; A29029.
DR      InterPro: IPR003322; Gag_P10.
DR      InterPro: IPR000721; Gag_P24.
DR      InterPro: IPR001878; ZnF_CCHC.
DR      Pfam; PF00098; zf-CCHC; 1.
DR      Pfam; PF00607; Gag_P24; 1.
DR      Pfam; PF02337; Gag_P10; 1.
DR      PRINTS; PR00939; CZHCZNFINGER.
DR      SMART; SM00343; ZnF_C2HC; 2.
DR      PROSITE; PS50158; ZF_CCHC; 1.
KW      Coat protein; Core protein; Nucleoprotein; Polyprotein; Myristate;
KW      Phosphorylation; DNA-binding; Zinc finger; Repeat.
FT      CHAIN          2               99       PHOSPHORYLATED PROTEIN PP21.
FT      CHAIN          100            195       PROTEIN P3.
FT      CHAIN          196            228       PROTEIN P8.
FT      CHAIN          229            252       MAJOR CORE PROTEIN P27.
FT      CHAIN          270            496       NUCLEIC ACID BINDING PROTEIN P14.
FT      CHAIN          497            591       CCHC-TYPE 1.
FT      2N_FING        525            542       CCHC-TYPE 2.
FT      2N_FING        552            569       CCHC-TYPE 2.
FT      LIPID          2               2       MYRISTATE.
SQ      SEQUENCE      591 AA; 65338 MW; 43833324EDDC467C CRC64;

Alignment Scores:
Pred. No.:              7.26e-06                Length:             591
Score:                  114.00                   Matches:             16
Percent Similarity:     50.50%                   Conservative:       35
Best Local Similarity:   15.84%                   Mismatches:        25
Query Match:            21.39%                   Indels:             25
DB:                      1                       Gaps:               4

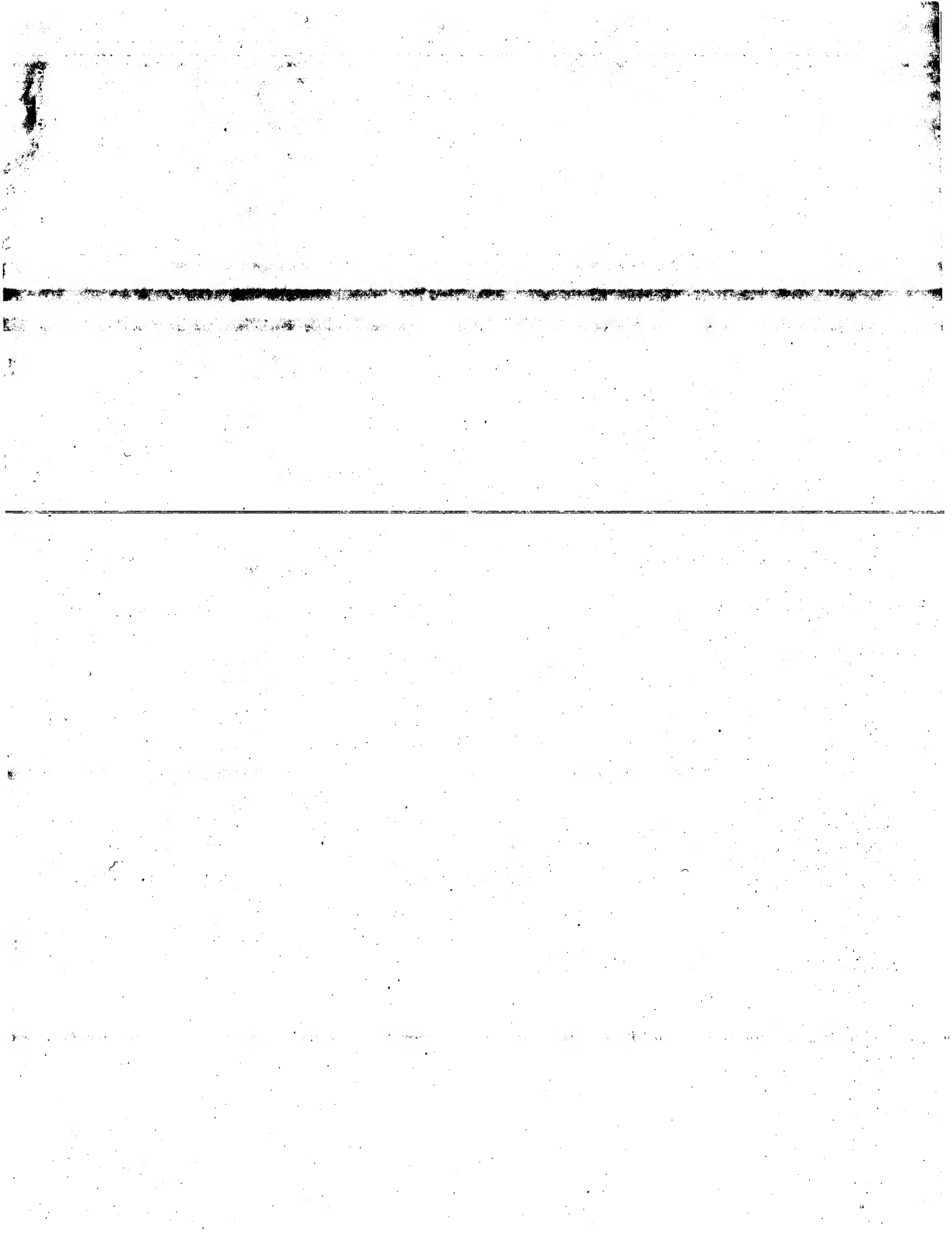
US-09-880-711-328 (1-350) x GAG-MMTVC (-1-591)
QY      .B4.GCTCTTCCTCCGCGCGCGCTTCACTACGCCCTTCATCATGCACGAGAGC-----1388
           |||:|||||:|||||         |||    |||         ::::|||
DB      404 Alalierproptroprogly-ValylslstyhValleualaelyleuysgncnglaangl 4422
QY      139 -----NCTGTATAC-----ATGCCTAG 155
           ::::::::::::::::::::NCTGTATAC-----ATGCCTAG 155
DB      423 ucluserlyrgluthrphelieserairgleugluvalaValtyrargyalmetprot 4433
QY      156 G-----NNNNCAACNCCTGACANNNTGGANAGACNCTNCAGCACATGNATGAT 2000
           |||         ::::::::::::::NNNNCAACNCCTGACANNNTGGANAGACNCTNCAGCACATGNATGAT 2000
DB      443 gelygluclyserrspilleuleullysglnleualaIripulsnsAlasnserrydecy 4633
QY      201 NCCTGCTNCCTGGGAATAGCCTGNCACAGATCAACACTACTGTCGACAGCANNAAGGAGCTT 2606
           |||         ::::::::::::::NCCTGCTNCCTGGGAATAGCCTGNCACAGATCAACACTACTGTCGACAGCANNAAGGAGCTT 2606
DB      463 sclnlspleulleatrgprometarlglysthgltYthrmetslinspyrrileatrglacy 4833
QY      261 NNAGACAGAAATGTATTNNNAACCNACTACTCTTTATNCCTTTTGCAANGANAACNTTCT 320
           ::::|||         ::::::::::::::NNAGACAGAAATGTATTNNNAACCNACTACTCTTTATNCCTTTTGCAANGANAACNTTCT 320
DB      483 sleusasp-----AlaserproalaValaValcInclgmetyralatyralaaIaIame 5000
QY      321 T 321
DB      500 t 500

```



Search completed: April 28, 2003, 15:36:36  
Job time : 17 secs

---



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame-plus.n2p model

Run on: April 28, 2003, 15:34:19 : Search time 36 Seconds

(without alignments)  
4006.472 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 533

Sequence: 1 atgagcgcacagctaccctgca.....ctgagagattcacaataaaa 350

Scoring table:

BLOSUM62DX  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 571580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_n2p\_model -DRV=xlp  
-O=/cgn2\_1/USPT0/spool/US09880711/runat\_28042003\_11:138\_855/app\_query.fasta\_1.519  
-DB=SPTRMBL\_21 -QPMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62dx -TRANS=human40.cdi  
-LIST=45 -DOCCALCN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZF=500 -MFILEN=0 -MAXLEN=2000000000  
-USER=US09880711.eccn.1.1.138.gnumat\_28042003.112136\_855 -NCPU=6 -ICPU=3  
-NO\_XLPPX -NO\_MKAP -LAKSEQUENT -NEG\_SCORES=0 -WAIT -DLOGLOG -DRV=TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mtc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	240	45.0	197	11	Q92100

2	240	45.0	220	11	Q9CPU5	Oryzops mus musculus
3	225.5	42.3	219	4	Q96J26	Oryzops mus musculus
4	125	23.5	478	8	Q8WJG7	Oryzops mus musculus
5	124	23.3	363	8	Q94X20	Oryzops mus musculus
6	124	24.5	431	8	P92380	Oryzops mus musculus
7	124	24.5	442	8	Q8WJG6	Oryzops mus musculus
8	124	24.1	512	8	Q9FKU2	Oryzops mus musculus
9	123	24.1	206	11	Q8VC52	Oryzops mus musculus
10	123	24.3	386	8	Q959C5	Oryzops mus musculus
11	123	24.3	478	8	Q8WJG7	Oryzops mus musculus
12	122	24.1	372	8	Q9XPM6	Oryzops mus musculus
13	122	22.9	409	8	Q9G5C9	Oryzops mus musculus
14	122	22.5	460	8	Q8WJG5	Oryzops mus musculus
15	122	24.1	505	8	Q19816	Oryzops mus musculus
16	122	24.1	505	8	Q8WJG1	Oryzops mus musculus
17	122	22.9	664	8	Q9JB53	Oryzops mus musculus
18	121	22.7	355	8	Q95FA6	Oryzops mus musculus
19	121	23.9	353	8	Q94X20	Oryzops mus musculus
20	121	23.7	405	8	Q8SK37	Oryzops mus musculus
21	121	22.7	505	8	Q8WJG1	Oryzops mus musculus
22	121	22.7	507	8	Q8SK39	Oryzops mus musculus
23	121	22.7	507	8	Q8SK38	Oryzops mus musculus
24	120	23.7	340	8	Q93556	Oryzops mus musculus
25	120	23.7	355	8	Q95FA6	Oryzops mus musculus
26	120	22.5	380	8	Q9G5C5	Oryzops mus musculus
27	120	22.5	439	8	Q33036	Oryzops mus musculus
28	120	22.5	442	8	Q8WJG5	Oryzops mus musculus
29	120	23.7	446	8	Q92138	Oryzops mus musculus
30	120	23.7	447	8	Q92635	Oryzops mus musculus
31	120	23.7	503	8	Q9TKA3	Oryzops mus musculus
32	120	23.5	508	8	Q8WJG0	Oryzops mus musculus
33	120	23.7	508	8	Q8WJG0	Oryzops mus musculus
34	119.5	22.4	515	8	Q9TKA3	Oryzops mus musculus
35	119	23.5	351	8	Q92832	Oryzops mus musculus
36	119	22.3	372	8	Q9XPM6	Oryzops mus musculus
37	119	22.3	380	8	Q9G5C5	Oryzops mus musculus
38	119	22.3	431	8	P92380	Oryzops mus musculus
39	119	22.3	436	8	Q32422	Oryzops mus musculus
40	119	22.3	447	8	Q32635	Oryzops mus musculus
41	119	22.5	447	8	Q8WJG8	Oryzops mus musculus
42	119	22.5	449	8	Q35544	Oryzops mus musculus
43	119	22.3	450	8	Q32803	Oryzops mus musculus
44	119	22.3	450	8	Q32575	Oryzops mus musculus
45	119	23.5	457	8	Q8WJG2	Oryzops mus musculus

## ALIGNMENTS

### RESULT 1

ID	Q92100	PRELIMINARY:	PR1:	197 AA.
AC	Q92100			
DT	01-DEC-2001 (TRENDEL, 19, Created)			
DT	01-DEC-2001 (TRENDEL, 19, Last sequence update)			
DT	01-JUN-2002 (TRENDEL, 21, Last annotation update)			
DE	Unknown (Protein for MCC118922).			
GN	RBPM5.			
OS	Mus musculus (Mouse).			
OC	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:			
OC	Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Mus.			
OX	NCBI_TaxID=10090:			
RP	111			
RA	Strasbourg R.			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC011284; AAI11284.1; -			
DR	MGD: MGI:1334446; Rbm5.			
DR	InterPro: IPR000504; RNA_rec_mot.			
DR	Pfam: PF00076; rrm; 1.			
DR	PROSITE: PS50102; RRM; 1.			
SO	SEQUENCE 197 AA: 21816 MW: 03457084F803CC78 CRC64:			

Alignment Scores:





[illegible]

RT on chloroplast trnK/matK sequences and its implications for  
 RT evolutionary patterns in Papilionoideae.";  
 RL Am. J. Bot. 87:418-430(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=1250829;  
 RA Lavin M., Pennington R.T., Kiltgaard B.B., Sprent J.I., de Lima H.C.,  
 "The dalbergioid legumes (Fabaceae): delimitation of a pantropical  
 monophyletic clade";  
 RT monophyletic clade";  
 RL Am. J. Bot. 88:503-533(2001).  
 CC -I- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP 11  
 CC -I- INTRONS (BY SIMILARITY).  
 CC -I- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 DR EMBL: AF142681: AAD52855.1: -;  
 DR InterPro: IPR000442: Intron\_maturase2.  
 DR InterPro: IPR002866: MatK\_N.  
 DR Pfam: PF01348: Intron\_maturase2; 1.  
 DR Pfam: PF01824: MatK\_N; 1.  
 DR Chloroplast; mRNA processing.  
 KW SEQUENCE 512 AA: 58638 MW: AA3BBI02F23A8D55 CRC64:

Alignment Scores:  
 Pred. No.: 3,16e-07 Length: 512  
 Score: 124.00 Matches: 11  
 Percent Similarity: 58.67% Conservative: 33  
 Best Local Similarity: 14.67% Mismatches: 12  
 Query Match: 24.51% Indels: 19  
 Dns: 8 Gaps: 2

US-09-880-711-328 (1-350) x Q9TKU2 (1-512)

OY 336 TGTCAAAANTTTCAGAGANNGTNTCTTTGCAAGAGNATAGGTAGTNGGTTNNTA 277  
 |||||  
 Db 91 SerGlnIleIleSerLeuGluGluAla-----GluLeuValIysSer 110  
 OY 276 ATACATTCGTCTCNNAAGCTTCCTTNGTGTGACAGTACGATCTGTCAGCGN 217  
 |||||  
 Db 111 SerSerSerLeuGluGluAla-----GluLeuValIysSer 122  
 OY 216 CATTCGCCAGNCAGCG-----NATCATATCATGT 190  
 |||||  
 Db 123 TyrAsnAsnLeuArgSerIleHisSerIlePheProPhePheGluAspIysPheThrTyr 142  
 OY 189 GGTGAGAGAGTTCAGANNNGTCAGGNGTGNNNNCTAGGCATGTA 145  
 |||||  
 Db 143 LeuAsnTyrValSerAspValArgIleProTyrProIleHisLeu 157

RESULT 9

ID O8VC52 PRELIMINARY: PRT: 206 AA.  
 AC O8VC52;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 22.5 kDa protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Scleromachi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=KIDNEY;  
 RA Strausberg K.;  
 RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: BC021788: AAI21788.1: -;  
 DR InterPro: IPR000504: RNA\_rec\_mot.  
 DR Pfam: PF00076: rrm; 1.  
 DR SMART: SM00360: RRM; 1.  
 DR PROSITE: PS0102: RRM; 1.  
 KW Hypothetical protein.

SO SEQUENCE 206 AA: 22462 MW: E90703F9B97ACD2 CRC64;

Alignment Scores:  
 Pred. No.: 4.27e-07 Length: 206  
 Score: 123.00 Matches: 30  
 Percent Similarity: 56.34% Conservative: 10  
 Best Local Similarity: 42.25% Mismatches: 21  
 Query Match: 23.08% Indels: 10  
 Dns: 11 Gaps: 2

US-09-880-711-328 (1-350) x O8VC52 (1-206)

OY 3 GAGCTCACAGTACCTGCTGACCTTACCCAGTACCCCTGAAGTGTGGCCCGTACCTTG 62  
 |||||  
 Db 135 AspIleuMetGlyThrAlaLeuIleProAlaSerProGluAlaThrIleProTyrProIleu 154  
 OY 63 TACCCAGCGAGTACGAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 116  
 |||||  
 Db 155 TyrThrThrGluIleThrProAlaIle---SerHisThrThrPheThrTyrProAlaAla 173  
 OY 117 -----GCTTCATGATGATGCCAGAGCGNCGTATACATGGCTAG 155  
 |||||  
 Db 174 ThrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 193  
 OY 156 GNNNMCACMCTACACNNNNGACACTNCTNAG 188  
 |||||  
 Db 194 AspThrThrGluGluGluGlyThrPylsTyrArgGln 204

RESULT 10

ID Q959C6 PRELIMINARY: PRT: 386 AA.  
 AC Q959C6;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Cytochrome b (fragment).  
 GN CYTB.  
 OS Galaxias sp. 'teviot'.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Galaxiidae; Galaxias.  
 OX NCBI\_TaxID=126309;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Waters J.M., Wallis G.P.;  
 RT "Cladogenesis and the loss of the marine life history phase in  
 freshwater fishes (osmeriformes: Galaxiidae).";  
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
 CC -I- FUNCTION: COMPONENT OF THE URIDUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
 CC -I- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN (BY SIMILARITY).  
 CC -I- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
 CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
 DR EMBL: AF267374: AAK56475.1: -;  
 DR InterPro: IPR00179: Cyt\_b\_b6.  
 DR Pfam: PF00032: cytochrome\_b\_c1; 1.  
 DR Pfam: PF00033: cytochrome\_b\_n; 1.  
 DR PROSITE: PS00193: CYTOCHROME\_B\_Q0; UNKNOWN 1.  
 KW electron transport; Heme; Mitochondrion; Respiratory chain;  
 KW Transmembrane; 1  
 FT NON\_TER 386  
 FT NON\_TER 386  
 SQ SEQUENCE 386 AA: 42757 MW: 39E3CB1D89E76394 CRC64:

Alignment Scores:  
 Pred. No.: 4.25e-07 Length: 386  
 Score: 123.00 Matches: 17  
 Percent Similarity: 54.55% Conservative: 25





Db 148 AsplleatgllleProtyrProtlehlsleu 157

RESULT 13

09G5C9 PRELIMINARY: PRT: 409 AA.

AC 09G5C9:

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).

GN COI.

OS Melanoplus truchensis.

OC Mitochondrion.

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;

OC Acridomorpha; Acridoidea; Acrididae; Melanoplineae; Melanoplus.

OX NCBI\_TaxID=12140;

RN [1]

RP SEQUENCE FROM N.A.

RA Knowles L., Olte D.:

RT "Phylogenetic analysis of montane grasshoppers from western North America (genus Melanoplus, Acrididae; Melanoplineae).";

KL Ann. Entomol. Soc. Amer. 93:421-431(2000).

CC -I- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).

CC -I- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME C + 2 H(2)O.

CC -I- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).

CC -I- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).

CC -I- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

DR EMBL: AF229005; AAC35281.1; -.

DR InterPro: IPR000883; COX1.

DR Pfam: PF00115; COX1.2.

DR PRINTS: PR01165; CYCOXIDASE1.

KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;

KW Respiratory chain; Transmembrane.

FT NON\_TER 1 1

FT NON\_TER 409 409

SO SEQUENCE 409 AA; 45084 MW; 0D177A0556CC655 CRC64;

Alignment Scores:

Pred. No.: 5.69e-07 Length: 409

Score: 122.00 Matches: 12

Percent Similarity: 54.22% Conservative: 33

Best Local Similarity: 14.46% Mismatches: 23

Query Match: 22.89% Indels: 15

Gaps: 3

DB: 8

US-09-880-711-328 (1-350) x 09G5C9 (1-409)

QY 124 TGCATGCCACAGCGCGGTATACATGCTAGGNNNNCCAGCNCCTGACNNNTGCANACTNC 183

Db 165 CysgIngluSerGlyLysIle-----GluSerPheGlyThrLeuGlyMetIle 180

QY 184 TNCACACATGATGATGATGCTGCGGAATGCGCTGNCAGATCAACTACTGAG 243

Db 181 TyrAlaMetLeuSerIleGlyLeuMetGlyPheIleValITrPalaHisIleMetPhe--- 199

QY 244 ACCANCAAGGAGCTTNGAGACAGA-----ATGTAT 276

Db 200 --ThrValGlyMetAspValAspThrArgAla\*\*\*\*\*AlaThrIleIleIle 218

QY 277 TANNACGACACTTATNCTTTGCAANGAANAACNTTCTTGANAATNTTGACA 336

Db 219 AlaValProthGlyIleLysValPheSerTrpLeuAlaThrLeuTrpGlyThrLysPhe 238

QY 337 GATTTCAGAT 345

Db 239 LysPheAsn 241

RESULT 14

08WJF5 PRELIMINARY: PRT: 460 AA.

AC 08WJF5:

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE ATP synthase beta subunit (Fragment).

GN ATPB.

OS Thissmia rodwayi.

OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Dioscoreales; Thismiaceae;

OC Thissmia.

OX NCBI\_TaxID=147078;

RN [1]

RP SEQUENCE FROM N.A.

RA Caddick L.R., Rudall P.J., Wilkin P., Hedderson T.A.J., Chase M.W.:

RT "Phylogeny and circumscription of Dioscoreales based on combined analyses of morphological and molecular data.";

RT submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF308034; AAL37109.1; -.

DR InterPro: IPR000793; ATPase\_a/b/c.

DR InterPro: IPR000154; ATPase\_a/b/cenile.

DR InterPro: IPR004100; ATPase\_a/bN.

DR Pfam: PF00009; ATP-synt\_ab\_1.

DR Pfam: PF0306; ATP-synt\_ab\_C\_1.

DR Pfam: PF02874; ATP-synt\_ab\_N\_1.

KW Chloroplast.

FT NON\_TER 1 1

FT NON\_TER 460 460

SO SEQUENCE 460 AA; 49784 MW; 025797491C7F9ECE CRC64;

Alignment Scores:

Pred. No.: 5.68e-07 Length: 460

Score: 122.00 Matches: 12

Percent Similarity: 61.54% Conservative: 36

Best Local Similarity: 15.38% Mismatches: 24

Query Match: 22.89% Indels: 6

Gaps: 2

DB: 8

US-09-880-711-328 (1-350) x 08WJF5 (1-460)

QY 133 AGAGGCTGTATATACATGCTAGGNNNNCCAGCNCCTGACNNNTGCANACTNCAGACA 192

Db 62 ArgGlyMetGluValIleAspThrGlyAlaProLeuSerValProValGlyGlyAlaThr 81

QY 193 -----TGNATGATGATGCTGCTGCGGAATGCGCTGNCAGATCAACTACTGAGACC 246

Db 82 LongIatArgIlePheAspValIleGlyIleProValAspAsnLeuGlyProValAspThr 101

QY 247 ANCAAGGAGACTTNGAGACAGA-----ATGTATTANNAACNACTACTTGA 254

Db 102 ArgThrThrSerProIleHisArgSerAlaProAlaPheIleGlnIleuAspThrLysLeu 121

QY 295 TNCCTTTGCAANGAANAACNTTCTTGANAATNTTGACAGATTGCAATTA 348

Db 122 SerIlePheGluThrGlyIleLysValValAspLeuAlaProTyrArgArg 139

RESULT 15

019816 PRELIMINARY: PRT: 505 AA.

AC 019816:

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Probable Intron maturase (Maturase K).

GN Allamanda cathartica (Yellow allamanda).

[illegible]

Search completed: April 28, 2003, 15:37:52  
Job time : 39 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OH nucleic - protein search, using frame\_plus\_n2p model

Run on: April 28, 2003, 15:35:24 : Search time 14 Seconds  
(without alignments)  
1471.146 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 533

Sequence: 1 atgagctcacagtactctca.....lttgcagatcttcaatataaa 350

Scoring table:

BLOSUM62DX  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Pgapop 5.0 , Pgapext 7.0  
Delop 5.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame\_n2p\_model -DEV=xlp  
-O=/cgn2\_1/USPTC/spool/US09880711/runat\_28042003\_112140\_888/app\_query.fasta.1.519  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=n2p.fai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62dx -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09880711.ecgn.1.1.28.ernat.28042003.112140.888 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -P\_GAPOP=6 -PGAPEXT=7  
-Y\_GAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Issued\_Patents\_AA:\*  
2: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backlist1.pep:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	131	25.9	400	4	US-09-076-137-42 Sequence 42, Appl
C 2	129	24.2	731	1	US-08-158-232-39 Sequence 39, Appl
C 3	125	25.5	731	1	US-08-158-232-39 Sequence 39, Appl
C 4	125	24.2	731	1	US-08-304-626-39 Sequence 39, Appl
C 5	129	25.5	731	1	US-08-304-626-39 Sequence 39, Appl
C 6	129	24.2	731	2	US-08-611-928-39 Sequence 39, Appl
C 7	129	25.5	731	2	US-08-611-928-39 Sequence 39, Appl
C 8	129	24.2	731	3	US-09-173-891-39 Sequence 39, Appl
C 9	125	25.5	731	3	US-09-173-891-39 Sequence 39, Appl
C 10	128	25.3	411	2	US-08-568-459A-19 Sequence 19, Appl
C 11	128	25.3	411	2	US-08-487-826B-31 Sequence 31, Appl
C 12	128	25.3	411	4	US-09-210-288-19 Sequence 19, Appl

13	128	24.0	716	4	US-09-076-137-41 Sequence 41, Appl
14	127	23.8	716	1	US-08-316-301A-41 Sequence 41, Appl
C 15	127	25.1	716	1	US-08-316-301A-41 Sequence 41, Appl
C 16	127	25.1	1040	4	US-08-974-549A-727 Sequence 727, App
C 17	126	24.9	401	1	US-08-316-301A-42 Sequence 42, Appl
C 18	126	24.9	716	4	US-09-076-137-41 Sequence 42, Appl
C 19	125	23.5	401	1	US-08-316-301A-42 Sequence 42, Appl
C 20	124	24.5	181	2	US-08-558-823-3 Sequence 3, Appl
C 21	124	24.3	400	4	US-09-076-137-42 Sequence 42, Appl
C 22	123.5	24.4	137	1	US-08-291-060B-2 Sequence 2, Appl
C 23	123	24.3	409	4	US-09-207-359B-46 Sequence 46, Appl
C 24	122	24.1	116	4	US-05-457-056-3 Sequence 3, Appl
C 25	122	22.9	1040	4	US-08-974-549A-727 Sequence 727, App
C 26	121	22.7	181	2	US-08-558-823-3 Sequence 3, Appl
C 27	121	22.7	409	4	US-09-207-359B-46 Sequence 46, Appl
C 28	121	22.7	461	1	US-08-196-989B-13 Sequence 13, Appl
C 29	121	22.7	451	2	US-08-760-936-13 Sequence 13, Appl
C 30	120	22.5	137	1	US-08-291-060B-2 Sequence 2, Appl
C 31	120	23.7	362	2	US-08-568-459A-18 Sequence 18, Appl
C 32	120	23.7	362	2	US-08-487-826B-30 Sequence 30, Appl
C 33	120	23.7	362	4	US-09-210-288-18 Sequence 18, Appl
C 34	120	23.7	411	2	US-08-568-459A-20 Sequence 20, Appl
C 35	120	23.7	411	2	US-08-487-826B-32 Sequence 32, Appl
C 36	120	23.7	411	4	US-09-210-288-20 Sequence 20, Appl
C 37	119	22.3	122	1	US-08-291-060B-3 Sequence 3, Appl
C 38	119	22.3	421	1	US-08-196-989B-12 Sequence 12, Appl
C 39	119	22.3	421	2	US-08-760-936-12 Sequence 12, Appl
C 40	119	23.5	820	4	US-09-391-104-7 Sequence 12, Appl
C 41	118	22.1	450	1	US-08-196-989B-11 Sequence 11, Appl
C 42	118	22.1	450	2	US-08-760-936-11 Sequence 11, Appl
C 43	118	23.3	461	1	US-08-196-989B-13 Sequence 13, Appl
C 44	118	23.3	461	2	US-08-760-936-13 Sequence 13, Appl
C 45	118	22.1	820	4	US-09-391-104-7 Sequence 7, Appl

#### ALIGNMENTS

RESULT 1  
US-09-076-137-42  
Sequence 42, Application US/09076137B

Patent No. 6166195

GENERAL INFORMATION:

APPLICANT: Schuebl, Harry E.

APPLICANT: Schuebl, George E.

APPLICANT: Payne, Jewel M.

APPLICANT: Narva, Kenneth E.

APPLICANT: Poncetrada, Luis

TITLE OF INVENTION: No. 6166195el Nematode-Active Toxins and Genes which Code

FILE REFERENCE: MA-20CC02

CURRENT APPLICATION NUMBER: US/09/076.137B

CURRENT FILING DATE: 1998-05-12

EARLIER APPLICATION NUMBER: 08/316.301

EARLIER FILING DATE: 1994-09-30

NUMBER OF SEQ ID NOS: 42

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 42

LENGTH: 400

TYPE: PRT

ORGANISM: Generic Formula 11

US-09-076-137-42

Alignment Scores:

Pred. No.: 117e-08

Percent Similarity: 131.00

Best Local Similarity: 100.008

Query Match: 83.33%

DBs: 25.89%

US-09-880-711-328 (1-350) x US-09-076-137-42 (1-400)

Qy 155 GNGTCNNNCCTAGGCAT 148

Length: 400  
Matches: 5  
Conservative: 1  
Mismatch: 0  
Indels: 0  
Gaps: 0

Db 7 |||||  
\*\*\*\*\*Prolyshis 12

## RESULT 2

US-08-158-232-39

Sequence 39, Application US/08158232  
Patent No. 5596071

## GENERAL INFORMATION:

APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Uick, Heidi Jane  
APPLICANT: Foncerrada, Luis  
APPLICANT: Schnepl, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/158,232  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 731 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: YES

US-08-158-232-39

## Alignment Scores:

Pred. No.:	2.15e-08	Length:	731
Score:	129.00	Matches:	2
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	24.20%	Indels:	0
DB:	1	Gaps:	0

US-09-880-711-328 (1-350) x US-08-158-232-39 (1-731)

QY 24 TACCC 29

Db 8 |||||  
8 TyPto 9

## RESULT 3

US-08-158-232-39

Sequence 39, Application US/08158232  
Patent No. 5596071

## GENERAL INFORMATION:

APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Uick, Heidi Jane  
APPLICANT: Foncerrada, Luis  
APPLICANT: Schnepl, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/158,232  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 731 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: YES

US-08-158-232-39

## Alignment Scores:

Pred. No.:	2.15e-08	Length:	731
Score:	129.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	25.49%	Indels:	0
DB:	1	Gaps:	0

US-09-880-711-328 (1-350) x US-08-158-232-39 (1-731)

QY 217 NCATCCGAGNG 206

```
||||:|||||
Db      7 ***TyPro*** 10

RESULT 4
US-08-304-626-39
: Sequence 39, Application US/08304626
: Patent No. 5616495
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel M.
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
: APPLICANT: Uick, Heidi Jane
: APPLICANT: Poncerrada, Luis
: APPLICANT: Schnepf, Harry E.
: APPLICANT: Schwab, George E.
: TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates
: TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding
: TITLE OF INVENTION: Hymenopteran-Active Toxins
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Salimanchik
: STREET: 2121 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32609
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/304,626
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/887,980
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Salimanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: M/SCJ 104
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 39:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 741 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: YES
: US-08-304-626-39

Alignment Scores:
Pred. No.: 2,15e-08 Length: 731
Score: 129.00 Matches: 2
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.20% Indels: 0
Gaps: 0
DB: 1

US-09-880-711-328 (1-350) x US-08-304-626-39 (1-731)

QY      24 1ACCCG 29
|||||
Db      8 TyPro 5

RESULT 5
US-08-304-626-39
: Sequence 39, Application US/08304626
```

```
: Patent No. 5616495
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel M.
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
: APPLICANT: Uick, Heidi Jane
: APPLICANT: Poncerrada, Luis
: APPLICANT: Schnepf, Harry E.
: APPLICANT: Schwab, George E.
: TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates
: TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding
: TITLE OF INVENTION: Hymenopteran-Active Toxins
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Salimanchik
: STREET: 2121 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/304,626
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/887,980
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Salimanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: M/SCJ 104
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 39:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 741 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: YES
: US-08-304-626-39

Alignment Scores:
Pred. No.: 2,15e-08 Length: 731
Score: 129.00 Matches: 3
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 25.49% Indels: 0
Gaps: 0
DB: 1

US-09-880-711-328 (1-350) x US-08-304-626-39 (1-731)

QY      217 NCAAGCGAGAG 205
|||||
Db      7 ***TyPro*** 10

RESULT 6
US-08-611-928 39
: Sequence 39, Application US/08611928
: Patent No. 5624792
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
```

APPLICANT: Uick, Heidi Jane  
APPLICANT: Foncerra, Luis  
APPLICANT: Schaeff, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Toxins Active  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,928  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/158,232  
FILING DATE: 24-NOV-1993  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 731 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: YES  
US-08-611-928-39

Alignment Scores:  
Pred. No.: 2.15e-08 Length: 731  
Score: 129.00 Matches: 2  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 24.20% Indels: 0  
Gaps: 0  
DB: 2

US-09-880-711-328 (1-350) x US-08-611-928-39 (1-731)

QY 24 TACCC 29  
|||||  
Db 8 TyrPro 9

RESULT 7  
US-08-611-928-39  
Sequence 39, Application US/08611928  
Patent No. 5824792

GENERAL INFORMATION:  
APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Uick, Heidi Jane  
APPLICANT: Foncerra, Luis  
APPLICANT: Schaeff, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Toxins Active  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,928  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/158,232  
FILING DATE: 24-NOV-1993  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 731 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: YES  
US-08-611-928-39

Alignment Scores:  
Pred. No.: 2.15e-08 Length: 731  
Score: 129.00 Matches: 3  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 75.00% Mismatches: 0  
Query Match: 25.49% Indels: 0  
Gaps: 0  
DB: 2

US-09-880-711-328 (1-350) x US-08-611-928-39 (1-731)

QY 217 NCATCCGAG 206  
|||||  
Db 7 \*\*\*TyrPro\*\*\* 10

## RESULT 8

US-09-173-891-39

: Sequence 39, Application US/05173891

: Patent No. 6077937

: GENERAL INFORMATION:

: APPLICANT: Payne, Jewel

: APPLICANT: Kennedy, M. Keith

: APPLICANT: Randall, John Brooks

: APPLICANT: Meier, Henry

: APPLICANT: Dick, Heidi Jane

: APPLICANT: Foncecrada, Luis

: APPLICANT: Schaepl, H. Ernest

: APPLICANT: Schwab, George E.

: APPLICANT: Fu, Jenny

: TITLE OF INVENTION: No. 6077937el Bacillus thuringiensis Toxins Active

: TITLE OF INVENTION: Against Hymenopteran Pests

: NUMBER OF SEQUENCES: 51

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: David R. Saliwanchik

: STREET: 2421 N.W. 41st Street, Suite A-1

: CITY: Gainesville

: STATE: FL

: COUNTRY: USA

: ZIP: 32606

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: OPERATING SYSTEM: IBM PC compatible

: SOFTWARE: Patent Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/09/173.891

: FILING DATE:

: CLASSIFICATION:

: APPLICATION DATA:

: PRIORITY APPLICATION NUMBER: 08/158,232

: FILING DATE:

: APPLICATION NUMBER: US 07/887,980

: FILING DATE: 22-MAY-1992

: PRIORITY APPLICATION DATA:

: APPLICATION NUMBER: US 07/797,645

: FILING DATE: 25-NOV-1991

: PRIORITY APPLICATION DATA:

: APPLICATION NUMBER: US 07/703,977

: FILING DATE: 22-MAY-1991

: ATTORNEY/AGENT INFORMATION:

: NAME: Saliwanchik, David R.

: REGISTRATION NUMBER: 31,794

: REFERENCE/DOCKET NUMBER: M/SCJ104.C1

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 904-375-8100

: TELEFAX: 904-372-5800

: INFORMATION FOR SEQ ID NO: 39:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 731 amino acids

: TYPE: amino acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: peptide

: HYPOTHETICAL: YES

: US-09-173-891-39

: Alignment Scores:

: Pred. No.: 2,15e-08

: Score: 129.00

: Percent Similarity: 100.00%

: Best Local Similarity: 100.00%

: Query Match: 24.20%

: DB: 3

: Gaps: 0

: US-09-880-711-328 (1-350) x US-09-173-891-39 (1-731)

Db 8 tyrPro 9

RESULT 9

US-09-173-891-39

: Sequence 39, Application US/09173891

: Patent No. 6077937

: GENERAL INFORMATION:

: APPLICANT: Payne, Jewel

: APPLICANT: Kennedy, M. Keith

: APPLICANT: Randall, John Brooks

: APPLICANT: Meier, Henry

: APPLICANT: Dick, Heidi Jane

: APPLICANT: Foncecrada, Luis

: APPLICANT: Schaepl, H. Ernest

: APPLICANT: Schwab, George E.

: APPLICANT: Fu, Jenny

: TITLE OF INVENTION: No. 6077937el Bacillus thuringiensis Toxins Active

: TITLE OF INVENTION: Against Hymenopteran Pests

: NUMBER OF SEQUENCES: 51

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: David R. Saliwanchik

: STREET: 2421 N.W. 41st Street, Suite A-1

: CITY: Gainesville

: STATE: FL

: COUNTRY: USA

: ZIP: 32606

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: OPERATING SYSTEM: IBM PC compatible

: SOFTWARE: Patent Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/09/173.891

: FILING DATE:

: CLASSIFICATION:

: APPLICATION DATA:

: PRIORITY APPLICATION NUMBER: 08/158,232

: FILING DATE:

: APPLICATION NUMBER: US 07/887,980

: FILING DATE: 22-MAY-1992

: PRIORITY APPLICATION DATA:

: APPLICATION NUMBER: US 07/797,645

: FILING DATE: 25-NOV-1991

: PRIORITY APPLICATION DATA:

: APPLICATION NUMBER: US 07/703,977

: FILING DATE: 22-MAY-1991

: ATTORNEY/AGENT INFORMATION:

: NAME: Saliwanchik, David R.

: REGISTRATION NUMBER: 31,794

: REFERENCE/DOCKET NUMBER: M/SCJ104.C1

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 904-375-8100

: TELEFAX: 904-372-5800

: INFORMATION FOR SEQ ID NO: 39:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 731 amino acids

: TYPE: amino acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: peptide

: HYPOTHETICAL: YES

: US-09-173-891-39

: Alignment Scores:

: Pred. No.: 2,15e-08

: Score: 129.00

: Percent Similarity: 100.00%

: Best Local Similarity: 100.00%

: Query Match: 25.49%

: DB: 3

: Gaps: 0

: US-09-880-711-328 (1-350) x US-09-173-891-39 (1-731)

OY 217 NCATCCGANG 206  
Db 7 \*\*\*TyPro\*\*\* 10

## RESULT 10

US-08-568-459A-19  
Sequence 19, Application US/08568459A

Patent No. 5849306

## GENERAL INFORMATION:

APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/568,459A

FILING DATE: 07-DEC-1995  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH21.001CP1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:  
LENGTH: 411 amino acids

TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide

HYPOTHETICAL: NO  
ANTI-SENSE: NO

FRAGMENT TYPE: internal  
ORIGINAL SOURCE:

US-08-568-459A-19

## Alignment Scores:

Pred. No.:	2.8e-08	Length:	411
Score:	128.00	Matches:	5
Percent Similarity:	88.89%	Conservative:	3
Best Local Similarity:	55.56%	Mismatches:	1
Query Match:	25.30%	Indels:	0
DB:	2	Gaps:	0

US-09-880-711-328 (1-350) x US-08-568-459A-19 (1-411)

OY 148 TGTATACAGNGCCTCTGGGATGCAGT 122

Db 278 CysIleSerCysLeuTyAlaCysasn 286

## RESULT 11:

US-08-487-826B-31  
Sequence 31, Application US/08487826B  
Patent No. 5993827

## GENERAL INFORMATION:

APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B

FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH21.001CP1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:  
LENGTH: 411 amino acids

TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide

HYPOTHETICAL: NO  
ANTI-SENSE: NO

FRAGMENT TYPE: internal  
ORIGINAL SOURCE:

US-08-487-826B-31

## Alignment Scores:

Pred. No.:	2.8e-08	Length:	411
Score:	128.00	Matches:	5
Percent Similarity:	88.89%	Conservative:	3
Best Local Similarity:	55.56%	Mismatches:	1
Query Match:	25.30%	Indels:	0
DB:	2	Gaps:	0

US-09-880-711-328 (1-350) x US-08-487-826B-31 (1-411)

OY 148 TGTATACAGNGCCTCTGGGATGCAGT 122

Db 278 CysIleSerCysLeuTyAlaCysasn 286

## RESULT 12

US-09-210-288-19  
Sequence 19, Application US/09210288

Patent No. 6392026

## GENERAL INFORMATION:

APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS



NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knoble Martens Olson & Neat  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/210,288  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fuller, Michael  
REGISTRATION NUMBER: 36,516  
REFERENCE/DOCKET NUMBER: NIH121.1PMDV1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 411 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
US-09-210-288-19

Alignment Scores:  
Pred. No.: 2.8e-08 Length: 411  
Score: 128.00 Matches: 5  
Percent Similarity: 88.89% Conservative: 3  
Best Local Similarity: 55.56% Mismatches: 1  
Query Match: 25.30% Indels: 0  
DB: 4 Gaps: 0

US-09-880-711-323 (1-350) x US-09-210-288-19 (1-411)

QY 148 TGTATACAGNCCCTGTGGCATGCACT 122  
||||| :||||:|||||:  
DB 278 cysliesercysleutyralacysasn 286

RESULT 13  
US-09-076-137-41  
Sequence 41, Application US/09076137B  
Patent No. 6163195  
GENERAL INFORMATION:  
APPLICANT: Schmepl, Harry E.  
APPLICANT: Scwab, George E.  
APPLICANT: Payne, Jewel M.  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Poncerada, Luis  
TITLE OF INVENTION: No. 6165195e1 Nematode-Active Toxins and Genes Which Code  
TITLE OF INVENTION: Therefor  
FILE REFERENCE: MA-20CCCD2  
CURRENT APPLICATION NUMBER: US/09/076,137B  
CURRENT FILING DATE: 1994-05-12  
EARLIER APPLICATION NUMBER: 08/316,301  
EARLIER FILING DATE: 1994-09-30  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 41  
LENGTH: 716  
TYPE: PRT

ORGANISM: Generic Formula 1  
US-09-076-137-41

Alignment Scores:  
Pred. No.: 2.86e-08 Length: 716  
Score: 128.00 Matches: 2  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 56.67% Mismatches: 0  
Query Match: 24.02% Indels: 0  
DB: 4 Gaps: 0

US-09-880-711-328 (1-350) x US-09-076-137-41 (1-716)

QY 98 TGTCCGCTT 106  
|||||:  
DB 622 CysArgTyr 624

RESULT 14  
US-08-316-301A-41  
Sequence 41, Application US/08316301A  
Patent No. 5753492  
GENERAL INFORMATION:  
APPLICANT: Schmepl, Harry E.  
APPLICANT: Schwab, George E.  
APPLICANT: Payne, Jewel M.  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Poncerada, Luis  
TITLE OF INVENTION: No. 5753492e1 Nematode-Active Toxins and Genes  
TITLE OF INVENTION: Which Code Therefor  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,301A  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/871,510  
FILING DATE: 23-APR-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/693,018  
FILING DATE: 03-MAY-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/565,544  
FILING DATE: 10-AUG-1990  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/084,653  
FILING DATE: 12-AUG-1987  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/830,050  
FILING DATE: 31-JAN-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lloyd, Jeff  
REGISTRATION NUMBER: 35,589  
REFERENCE/DOCKET NUMBER: MA20CCCD1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 41:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 716 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-316-301A-41  
  
 Alignment Scores:  
 Pred. No.: 3,82e-08 Length: 716  
 Score: 127.00 Matches: 2  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 66.67% Mismatches: 0  
 Query Match: 23.83% Indels: 0  
 DB: 1 Gaps: 0  
  
 US-09-880-711-328 (1-350) x US-08-316-301A-41 (1-716).  
 QY 98 TGGCGCTT 106  
 |||||:  
 Db 622 CysArgTyr 624  
  
 RESULT 15  
 US-08-316-301A-41  
 Sequence 41, Application.US/08316301A  
 Patent No. 5753492  
 GENERAL INFORMATION:  
 APPLICANT: Schmepl, Harry E.  
 APPLICANT: Schwab, George E.  
 APPLICANT: Payne, Jewel M.  
 APPLICANT: Narva, Kenneth E.  
 APPLICANT: Fonzerrada, Luis  
 TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes  
 TITLE OF INVENTION: Which Code Therefor  
 NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Saliwanchik & Saliwanchik  
 STREET: 2421 N.W. 41st Street, Suite A-1  
 CITY: Gainesville  
 STATE: FL  
 COUNTRY: USA  
 ZIP: 32606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/316,301A  
 FILING DATE: 30-SEP-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/871,510  
 FILING DATE: 23-APR-1992  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/693,018  
 FILING DATE: 03-MAY-1991  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/565,544  
 FILING DATE: 10-AUG-1990  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/084,653  
 FILING DATE: 12-AUG-1987  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/830,050  
 FILING DATE: 31-JAN-1992  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:

NAME: Lloyd, Jeff  
 REGISTRATION NUMBER: 35,589  
 REFERENCE/DOCKET NUMBER: MA20CCCD1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 904-375-8100  
 TELEFAX: 904-372-5800  
 INFORMATION FOR SEQ ID NO: 41:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 716 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-316-301A-41  
  
 Alignment Scores:  
 Pred. No.: 3,82e-08 Length: 716  
 Score: 127.00 Matches: 3  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 75.00% Mismatches: 0  
 Query Match: 25.10% Indels: 0  
 DB: 1 Gaps: 0  
  
 US-09-880-711-328 (1-350) x US-08-316-301A-41 (1-716)  
 QY 23 AAGTGCAGGTAC 12  
 ::::|  
 Db 621 ArgCysArgTyr 624

Search completed: April 28, 2003, 15:39:12  
 Job time : 17 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OH nucleic - protein search, using frame\_plus\_n2p model

Run on: April 28, 2003, 15:39:10 ; Search time 19 Seconds

(without alignments)  
2552.151 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 533  
Sequence: 1 atgagctacagctacgta.....ltgagagatctcaataaa 350

Scoring table:

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delpop 6.0 , Delpext 7.0

Searched: 301932 seqs, 80125603 residues

Total number of hits satisfying chosen parameters: 603864

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0K  
Maximum Match 100K  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame-n2p.model -DPV=xlp  
-CGN2\_L1/USPT0.SPOOL/US09880711/runat.28042003\_112140\_504/app\_query.fasta.1.519  
-DB=publshed\_Applications\_AA -QPMT=fastan -SUFEX=n2p.rapb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blissum62px  
-TRANS=human40.cgi -LIST=45 -DLOCALIGN=200 -THR\_SCORE=PCT -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09880711\_6CGN\_1.1.17\_6runat.28042003\_112140\_504  
-KCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NRC\_SCORES=0 -NATV -LONGLOC  
-DEV\_T1MOUT=120 -MARN\_T1MOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FgapEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications\_AA:\*

1:	/cgcn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2:	/cgcn2_6/ptodata/1/pubpaa/PCT_MIN_PUB.pep.*
3:	/cgcn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4:	/cgcn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cgcn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6:	/cgcn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7:	/cgcn2_6/ptodata/1/pubpaa/PCNUS_PUBCOMB.pep.*
8:	/cgcn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cgcn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10:	/cgcn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11:	/cgcn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12:	/cgcn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13:	/cgcn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14:	/cgcn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	132.5	24.9	163	10	US-09-867-550-834
2	128	24.0	232	10	US-09-854-122-1
3	128	25.3	411	9	US-10-153-273-19
4	126	23.6	1208	9	US-10-101-812-11

C 5	126	24.9	1208	9	US-10-101-812-11	Sequence 11, Appl
C 6	125	24.7	513	9	US-09-051-013-2	Sequence 2, Appl1
C 7	125	24.7	1206	9	US-10-101-812-9	Sequence 2, Appl1
C 8	125	24.7	1206	9	US-10-101-662A-14	Sequence 14, Appl
C 9	124	23.3	175	10	US-09-795-338A-26	Sequence 26, Appl
C 10	124	23.3	413	9	US-10-154-155-1	Sequence 1, Appl1
C 11	124	24.3	513	9	US-09-051-013-2	Sequence 2, Appl1
C 12	124	24.5	1210	9	US-10-101-812-6	Sequence 5, Appl1
C 13	124	24.5	1210	9	US-10-101-652A-13	Sequence 13, Appl
C 14	123	24.3	128	10	US-09-908-805B-59	Sequence 59, Appl
C 15	123	24.3	175	10	US-09-795-338A-26	Sequence 26, Appl
C 16	123	24.1	261	10	US-09-916-790-37	Sequence 37, Appl
C 17	122	24.1	1206	9	US-10-139-583-3	Sequence 3, Appl1
C 18	121	22.7	1206	9	US-10-101-812-9	Sequence 9, Appl1
C 19	121	22.7	1206	9	US-10-101-652A-14	Sequence 14, Appl
C 20	120	22.5	128	10	US-09-908-805B-59	Sequence 59, Appl
C 21	120	23.7	362	10	US-10-153-273-18	Sequence 18, Appl
C 22	120	23.7	364	10	US-09-051-755-18	Sequence 18, Appl
C 23	120	23.7	411	9	US-10-153-273-20	Sequence 20, Appl
C 24	120	22.5	1208	9	US-10-101-662A-12	Sequence 12, Appl
C 25	120	23.7	1208	9	US-10-101-652A-12	Sequence 12, Appl
C 26	120	22.5	1333	9	US-09-972-115A-2	Sequence 2, Appl1
C 27	119	23.5	261	10	US-09-916-790-37	Sequence 37, Appl
C 28	119	23.3	270	10	US-09-815-915-17	Sequence 17, Appl
C 29	119	23.5	270	10	US-09-815-915-17	Sequence 17, Appl
C 30	119	22.3	270	10	US-09-910-150-37	Sequence 37, Appl
C 31	119	23.5	270	10	US-09-910-150-37	Sequence 37, Appl
C 32	116	22.3	270	10	US-09-834-456A-6	Sequence 6, Appl1
C 33	119	23.5	270	10	US-09-834-456A-6	Sequence 6, Appl1
C 34	119	22.3	270	10	US-09-515-806-5	Sequence 5, Appl1
C 35	115	23.5	270	10	US-09-515-806-5	Sequence 5, Appl1
C 36	115	22.3	270	10	US-09-860-352A-4	Sequence 4, Appl1
C 37	119	22.3	1034	10	US-09-858-754-5	Sequence 5, Appl1
C 38	116	22.3	1267	9	US-09-972-115A-4	Sequence 4, Appl1
C 39	119	23.5	1368	9	US-10-029-413A-10	Sequence 10, Appl
C 40	118	22.1	1210	9	US-10-101-812-6	Sequence 6, Appl1
C 41	118	22.1	1210	9	US-10-101-662A-13	Sequence 13, Appl
C 42	117	23.1	232	10	US-09-854-122-1	Sequence 1, Appl1
C 43	117	23.1	270	10	US-09-860-352A-4	Sequence 4, Appl1
C 44	117	22.0	364	10	US-09-051-755-18	Sequence 18, Appl
C 45	117	23.1	413	9	US-10-154-155-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-09-867-550-834  
Sequence 834, Application US/09867550  
Patent No. US20020082206A1  
GENERAL INFORMATION:  
APPLICANT: Leach, Martin D.  
APPLICANT: Mehrtz, Rued,  
APPLICANT: Conley, Pamela  
APPLICANT: Law, Debbie  
TITLE OF INVENTION: Topper, James  
TITLE OF INVENTION: Thereby  
FILE REFERENCE: 21402-013  
CURRENT APPLICATION NUMBER: US/09/867,550  
CURRENT FILING DATE: 2001-09-20  
PRIOR APPLICATION NUMBER: HSN 60/208,427  
PRIOR FILING DATE: 2000-05-30  
NUMBER OF SEQ ID NOS: 2125  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 834  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-867-550-834  
Alignment Scores: 1.33e-07 Length: 163  
Pred. No.: 132.50 Matches: 29

Percent Similarity: 66.00%  
Best Local Similarity: 58.00%  
Query Match: 24.86%  
DB: 10  
Gaps: 2

US-09-880-711-328 (1-350) x US-09-867-550-834 (1-163)

QY 3 GAGCTACAGTACCTGACCTTACCCAGTACCTGAGTGGCCCGCTACCTCTG 62  
DB 111 Aspleumetgylalaialaleuileprolaserproglualatrpalaiprotyleu 130  
QY 63 TACCCAGCGAGTACGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 116  
DB 131 Tyrthrtrhgltleuthrprolaile---serhisalaalaphethrtyrprothala 149  
QY 117 -----GCTTCAGTCGATGCGCCAG 134  
DB 150 ThrAlaAlaAlaAlaAlaAlaAlaAlaIn 159

## RESULT 2

US-09-854-122-1  
Sequence 1, Application US/09854122  
Patent No. US20020016980A1  
GENERAL INFORMATION:  
APPLICANT: ALBERTE, RANDALL S.  
TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
FILE REFERENCE: PHA-007.01  
CURRENT APPLICATION NUMBER: US/09/854,122  
CURRENT FILING DATE: 2001-09-10  
PRIORITY APPLICATION NUMBER: 60/202,529  
PRIORITY FILING DATE: 2000-05-10  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 232  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Consensus  
NAME/KEY: MOD\_RES  
LOCATION: (8)  
OTHER INFORMATION: N or T  
NAME/KEY: MOD\_RES  
LOCATION: (10)..(219)  
OTHER INFORMATION: Any amino acid; this region may vary in length from  
NAME/KEY: MOD\_RES  
LOCATION: (223)..(224)  
OTHER INFORMATION: Any amino acid  
NAME/KEY: MOD\_RES  
LOCATION: (229)..(230)  
OTHER INFORMATION: Any amino acid  
US-09-854-122-1

## Alignment Scores:

Pred. No.: 4,46e-07  
Score: 128.00  
Percent Similarity: 60.00%  
Best Local Similarity: 80.00%  
Query Match: 24.02%  
DB: 10  
Gaps: 0

US-09-880-711-328 (1-350) x US-09-854-122-1 (1-232)

QY 108 ACCCTACCCGCTTCA 122  
DB 1 ThrTyPProLysSer 5

## RESULT 3

US-10-153-273-19  
Sequence 19, Application US/10153273

Patent No. US20020169305A1

GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.

Chitnis, Chetan  
Miller, Louis H.  
Peterson, David S.  
Su, Xin-zhaun  
Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153,273  
FILING DATE: 21-May-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/210,288

FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael  
REGISTRATION NUMBER: 36,516

REFERENCE/DOCKET NUMBER: NIH121,1EWDV1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:

LENGTH: 411 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
HYPOTHETICAL: NO

ANTI-SENSE: NO  
FRAGMENT TYPE: Internal

ORIGINAL SOURCE:  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-10-153-273-19

## Alignment Scores:

Pred. No.: 4,37e-07  
Score: 128.00  
Percent Similarity: 88.89%  
Best Local Similarity: 55.56%  
Query Match: 25.30%  
DB: 9  
Gaps: 0

US-09-880-711-328 (1-350) x US-10-153-273-19 (1-411)

QY 148 TGTATACAGNGCCTCTGGGCAATGCAGT 122  
DB 278 CysIleSerCysLeuTyPAlaCysAsn 286

## RESULT 4

US-10-101-812-11  
Sequence 11, Application US/10101812  
Publication No. US20020192737A1  
GENERAL INFORMATION:  
APPLICANT: Kaelin Jr., William G.  
APPLICANT: Ivan, Mitcea  
TITLE OF INVENTION: Pharmaceuticals and Methods for Treating Hypoxia and

```

: TITLE OF INVENTION: Screening Methods Thereof
: FILE REFERENCE: 20363-014
: CURRENT APPLICATION NUMBER: US/10/101,812
: CURRENT FILING DATE: 2002-06-07
: PRIOR APPLICATION NUMBER: 60/277,425
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/277,431
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/277,440
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/332,493
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/332,334
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/345,200
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/345,131
: PRIOR FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/342,598
: PRIOR FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/345,132
: PRIOR FILING DATE: 2001-12-20
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 11
: LENGTH: 1208
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Consensus
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)..(600)
: OTHER INFORMATION: A peptide of 1 to 600 amino acids long wherein Xaa
: OTHER INFORMATION: Is any amino acid.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (601)
: OTHER INFORMATION: wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (603)
: OTHER INFORMATION: wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (605)
: OTHER INFORMATION: wherein Xaa is Ser, Thr, or Tyr.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (606)..(608)
: OTHER INFORMATION: wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: US-10-101-812-11

Alignment Scores:
Pred. No.: 7.27e-07 Length: 1208
Score: 126.00 Matches: 3
Percent Similarity: 100.00% Conservatve: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 23.64% Indels: 0
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-10-101-812-11 (1-1208):
OY 291 CTTATNCGT 299
Db 602 Leu***Pro 604

RESULT 5
US-10-101-812-11
```

```

: Sequence 11, Application US/10101812
: Publication No. US20020192737A1
: GENERAL INFORMATION:
: APPLICANT: Kaelin Jr., William G
: APPLICANT: Ivan, Mircea
: TITLE OF INVENTION: Pharmaceuticals and Methods for Treating Hypoxia and
: TITLE OF INVENTION: Screening Methods Thereof
: FILE REFERENCE: 20363-014
: CURRENT APPLICATION NUMBER: US/10/101,812
: CURRENT FILING DATE: 2002-06-07
: PRIOR APPLICATION NUMBER: 60/277,425
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/277,431
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/277,440
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/332,493
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/332,334
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/345,200
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/345,131
: PRIOR FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/342,598
: PRIOR FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/345,132
: PRIOR FILING DATE: 2001-12-20
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 11
: LENGTH: 1208
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Consensus
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)..(600)
: OTHER INFORMATION: A peptide of 1 to 600 amino acids long wherein Xaa
: OTHER INFORMATION: Is any amino acid.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (601)
: OTHER INFORMATION: wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (603)
: OTHER INFORMATION: wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (605)
: OTHER INFORMATION: wherein Xaa is Ser, Thr, or Tyr.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (606)..(608)
: OTHER INFORMATION: wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: US-10-101-812-11

Alignment Scores:
Pred. No.: 7.27e-07 Length: 1208
Score: 126.00 Matches: 4
Percent Similarity: 100.00% Conservatve: 1
Best local Similarity: 80.00% Mismatches: 0
Query Match: 24.90% Indels: 0
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-10-101-812-11 (1-1208)
```

```

TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
  OTHER INFORMATION: Description of Artificial Sequence: Consensus
  OTHER INFORMATION: Target Peptide
FEATURE:
  NAME/KEY: VARIANT
  LOCATION: (1)..(600)
  OTHER INFORMATION: A peptide from 1 to 600 amino acids long wherein
  OTHER INFORMATION: Xaa is any amino acid.
FEATURE:
  NAME/KEY: VARIANT
  LOCATION: (602)
  OTHER INFORMATION: Wherein Xaa is one or more of any of the amino
  OTHER INFORMATION: acids.
FEATURE:
  NAME/KEY: VARIANT
  LOCATION: (605)
  OTHER INFORMATION: Wherein Xaa is one or more of any of the amino
  OTHER INFORMATION: acids.
FEATURE:
  NAME/KEY: VARIANT
  LOCATION: (607)..(1206)
  OTHER INFORMATION: A peptide from 1 to 600 amino acids long wherein
  OTHER INFORMATION: Xaa is any amino acid.
US-10-101-812-9

Alignment Scores:
Pred: No.:          9.54e-07          Length: 1206
Score: 125.00       Matches: 3
Percent Similarity: 100.00%          Conservative: 1
Best Local Similarity: 75.00%         Mismatches: 0
Query Match: 24,708                  Indels: 0
Db: 9                                Gaps: 0

US-09-880-711-328 (1-350) x US-10-101-812-9 (1-1206)
Oy 335 CTCGAAATNT 324
Db 603 LeuLys***Leu 606

RESULT 8
US-10-101-662A-14
: Sequence 14, Application US/10101662A
: Publication No. US20030022198A1
GENERAL INFORMATION:
  APPLICANT: Kaelin Jr., William G
  APPLICANT: Livingston, David A
  APPLICANT: Klm, William
  TITLE OF INVENTION: Light Emitting Fusion Proteins and Diagnostic and
  FILE REFERENCE: 20363-009
  CURRENT APPLICATION NUMBER: US/10/101,662A
  CURRENT FILING DATE: 2002-03-19
  PRIOR APPLICATION NUMBER: 60/277,425
  PRIOR FILING DATE: 2001-03-20
  PRIOR APPLICATION NUMBER: 60/277,431
  PRIOR FILING DATE: 2001-03-20
  PRIOR APPLICATION NUMBER: 60/277,440
  PRIOR FILING DATE: 2001-03-20
  PRIOR APPLICATION NUMBER: 60/332,493
  PRIOR FILING DATE: 2001-11-09
  PRIOR APPLICATION NUMBER: 60/345,131
  PRIOR FILING DATE: 2001-12-20
  PRIOR APPLICATION NUMBER: 60/342,598
  PRIOR FILING DATE: 2001-12-20
  PRIOR APPLICATION NUMBER: 60/345,132
  PRIOR FILING DATE: 2001-12-20
  PRIOR APPLICATION NUMBER: 60/332,334
  PRIOR FILING DATE: 2001-11-09
  PRIOR APPLICATION NUMBER: 60/345,200
  PRIOR FILING DATE: 2001-11-09
  NUMBER OF SEQ ID NOS: 24

```

SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 14  
LENGTH: 1206  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (607)..(600)  
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is  
OTHER INFORMATION: any amino acid.  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (602)  
OTHER INFORMATION: wherein Xaa is any amino acid.  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (605)  
OTHER INFORMATION: wherein Xaa is any amino acid.  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (607)..(1206)  
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is  
OTHER INFORMATION: any amino acid.  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: consensus  
US-10-101-662A-14

Alignment Scores:  
Pred. No.: 9,54e-07 Length: 1206  
Score: 125.00 Matches: 3  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 75.00% Mismatches: 0  
Query Match: 24.70% Indels: 0  
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-10-101-662A-14 (1-1236)

QY 335 CTCAMANTNT 324

DB 603 LeuLys\*\*Leu 606

RESULT 9  
US-09-796-338A-26  
Sequence 26, Application US/09796338A  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc.  
TITLE OF INVENTION: 1983, 52881, 2398, 45449, 50289, AND  
TITLE OF INVENTION: 52872, NOVEL G PROTEIN-COUPLED RECEPTORS AND USES THEREFOR  
FILE REFERENCE: 10448-020001  
CURRENT APPLICATION NUMBER: US/09/796,338A  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/186,059  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 26  
LENGTH: 175  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: exemplary motif  
NAME/KEY: VARIANT  
LOCATION: 1-4, 6-53, 55-66, 68-137, 139-144, 146, 147, 150-170, 172,  
LOCATION: 173, 175  
OTHER INFORMATION: Xaa = any amino acid  
NAME/KEY: VARIANT  
LOCATION: 149  
OTHER INFORMATION: Xaa = any aromatic amino acid  
US-09-796-338A-26

Alignment Scores:

Pred. No.: 1.33e-06 Length: 175  
Score: 124.00 Matches: 2  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 23.25% Indels: 0  
DB: 10 Gaps: 0

US-09-880-711-328 (1-350) x US-09-796-338A-26 (1-175)

QY 138 CNETGT 143

DB 4 \*\*\*Cys 5

RESULT 10  
US-10-194-155-1  
Sequence 1, Application US/10194155  
Publication No. US20030045679A1  
GENERAL INFORMATION:  
APPLICANT: Crawford, Dana R  
TITLE OF INVENTION: Calcineurin Modulators  
FILE REFERENCE: 19705-005  
CURRENT APPLICATION NUMBER: US/10/194,155  
CURRENT FILING DATE: 2002-07-12  
PRIOR APPLICATION NUMBER: 60/305,202  
PRIOR FILING DATE: 2001-07-13  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 413  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Calcineurin  
OTHER INFORMATION: modulator  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)..(200)  
OTHER INFORMATION: wherein Xaa is any amino acid, or no amino acid,  
OTHER INFORMATION: such that Xaa 1-200 and Xaa 205-405 combine to  
OTHER INFORMATION: represent 200 amino acids total.  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (214)..(413)  
OTHER INFORMATION: wherein Xaa is any amino acid, or no amino acid,  
OTHER INFORMATION: such that Xaa 1-200 and Xaa 205-405 combine to  
OTHER INFORMATION: represent 200 amino acids total.  
US-10-194-155-1

Alignment Scores:  
Pred. No.: 1.3e-06 Length: 413  
Score: 124.00 Matches: 4  
Percent Similarity: 80.00% Conservative: 0  
Best Local Similarity: 80.00% Mismatches: 1  
Query Match: 23.25% Indels: 0  
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-10-154-155-1 (1-413)

QY 81 CTTGCTTCTGCT 95

DB 208 ProAlaserProPro 212

RESULT 11  
US-09-051-013-2  
Sequence 2, Application US/09051013  
Publication No. US20020188103A1  
GENERAL INFORMATION:  
APPLICANT: Baslor, Timothy H.  
TITLE OF INVENTION: CHIMERIC DNA-BINDING/DNA METHYLTRANSFERASE NUCLEIC ACID  
FILE REFERENCE: 48075/B/PCT/US  
CURRENT APPLICATION NUMBER: US/09/051,013  
CURRENT FILING DATE: 1998-10-09

NUMBER OF SEQ ID NOS: 8  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 2  
LENGTH: 513  
TYPE: PRT  
ORGANISM: Spiroplasma  
FEATURE:  
OTHER INFORMATION: X in this sequence is unknown  
US-09-051-013-2

Alignment Scores:  
Pred. No.: 129e-06 Length: 513  
Score: 124.00 Matches: 4  
Percent Similarity: 76.47% Conservative: 9  
Best Local Similarity: 23.53% Mismatches: 4  
Query Match: 23.26% Indels: 0  
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-09-051-013-2 (1-513)

OY 247 ANCNAGGAGCTTNGAGACAGATGTATTANNACNACCTTATTC 297  
DB 1 MetArgGlySerHisHisHisHisHisGlyIleCysThrThrMetSer 17

RESULT 12  
US-10-101-812-6  
Sequence 6, Application US/10101812.  
Publication No. US20020192737A1  
GENERAL INFORMATION:  
APPLICANT: Kaelin Jr., William G  
TITLE OF INVENTION: Pharmaceuticals and Methods for Treating Hypoxia and  
FILE REFERENCE: 20363-014  
CURRENT FILING DATE: 2002-06-07  
PRIOR FILING DATE: 2002-06-07  
PRIOR APPLICATION NUMBER: 60/277,425  
PRIOR FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: 60/277,431  
PRIOR FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: 60/277,440  
PRIOR FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: 60/332,493  
PRIOR FILING DATE: 2001-11-09  
PRIOR APPLICATION NUMBER: 60/332,334  
PRIOR FILING DATE: 2001-11-09  
PRIOR APPLICATION NUMBER: 60/345,200  
PRIOR FILING DATE: 2001-11-09  
PRIOR APPLICATION NUMBER: 60/345,131  
PRIOR FILING DATE: 2001-12-20  
PRIOR APPLICATION NUMBER: 60/342,598  
PRIOR FILING DATE: 2001-12-20  
PRIOR APPLICATION NUMBER: 60/345,132  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 6  
LENGTH: 1210  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Consensus  
OTHER INFORMATION: Target Peptide  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)..(600)  
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is  
OTHER INFORMATION: any amino acid.  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (611)..(1210)  
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is  
OTHER INFORMATION: any amino acid.

US-10-101-812-6

Alignment Scores:  
Pred. No.: 125e-06 Length: 1210  
Score: 124.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 0  
Query Match: 24.51% Indels: 0  
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-10-101-812-6 (1-1210)

OY 178 NTCCANNNGTCAGGNGTGNMNC 155  
DB 599 \*\*\*\*GlySerGlyIlePheLeu 606

RESULT 13  
US-10-101-662A-13  
Sequence 13, Application US/10101662A  
Publication No. US20030022198A1  
GENERAL INFORMATION:  
APPLICANT: Kaelin Jr., William G  
APPLICANT: Livingston, David A  
TITLE OF INVENTION: Light Emitting Fusion Proteins and Diagnostic and  
FILE REFERENCE: 20363-009  
CURRENT FILING DATE: 2002-03-19  
PRIOR FILING DATE: 2002-03-19  
PRIOR APPLICATION NUMBER: 60/277,425  
PRIOR FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: 60/277,431  
PRIOR FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: 60/277,440  
PRIOR FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: 60/332,493  
PRIOR FILING DATE: 2001-11-09  
PRIOR APPLICATION NUMBER: 60/345,131  
PRIOR FILING DATE: 2001-12-20  
PRIOR APPLICATION NUMBER: 60/342,598  
PRIOR FILING DATE: 2001-12-20  
PRIOR APPLICATION NUMBER: 60/345,132  
PRIOR FILING DATE: 2001-12-20  
PRIOR APPLICATION NUMBER: 60/332,334  
PRIOR FILING DATE: 2001-11-09  
PRIOR APPLICATION NUMBER: 60/345,200  
PRIOR FILING DATE: 2001-11-09  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 13  
LENGTH: 1210  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Consensus  
OTHER INFORMATION: Target Peptide  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)..(600)  
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is  
OTHER INFORMATION: any amino acid.  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (611)..(1210)  
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is  
OTHER INFORMATION: any amino acid.  
US-10-101-662A-13

Alignment Scores:  
Pred. No.: 1.25e-06 Length: 1210  
Score: 124.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 0



```

Query Match:          24.51%      Index:          0
DB:                   9           Gaps:          0

US-09-880-711-328 (1-350) x US-10-101-662A-13 (1-1210)

QY      178 NTCANNNGTCAGCGNGTGNMNC 155
      |||||:|||||:|||||:
      DB      599 *****GlySerGlyIlePheLeu 606

RESULT 14
US-09-908-805B-59
: Sequence 59, Application US/09908805B
: Patent No. US20020147307A1
: GENERAL INFORMATION:
: APPLICANT: HILTON, Douglas J
: APPLICANT: ALEXANDER, Warren S
: APPLICANT: VINCE, Elizabeth M
: APPLICANT: WILSON, Tracey A
: APPLICANT: RICHARDSON, Rachael T
: APPLICANT: STARR, Robyn
: APPLICANT: NICHOLSON, Sandra E
: APPLICANT: METCALF, Donald
: APPLICANT: NICOLA, Nicos A
: TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
: FILE REFERENCE: 109762
: CURRENT APPLICATION NUMBER: US/09/908.805B
: CURRENT FILING DATE: 2001-07-19
: PRIOR APPLICATION NUMBER: 08/962,560
: NUMBER OF SEQ ID NOS: 81
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 59
: LENGTH: 128
: TYPE: PRT
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: UNSURE
: LOCATION: (1)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE
: LOCATION: (2)
: OTHER INFORMATION: Xaa is any amino acid residue
: NAME/KEY: UNSURE
: LOCATION: (3)
: OTHER INFORMATION: Xaa is Pro, Thr or Ser
: NAME/KEY: UNSURE
: LOCATION: (4)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE
: LOCATION: (5)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (6)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (7)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala, Phe, Tyr or Trp
: NAME/KEY: UNSURE
: LOCATION: (8)
: OTHER INFORMATION: Xaa is Cys, Thr or Ser
: NAME/KEY: UNSURE
: LOCATION: (9)
: OTHER INFORMATION: Xaa is Arg, Lys or His
: NAME/KEY: UNSURE
: LOCATION: (10)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (11)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (12)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE

```

```

: LOCATION: (13)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (14)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (15)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (16)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala, Pro, Gly, Cys, Thr
: NAME/KEY: UNSURE
: LOCATION: (17)
: OTHER INFORMATION: Xaa can be any amino acid or no amino acid, position 17-65
: NAME/KEY: UNSURE
: LOCATION: (67)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE
: LOCATION: (68)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (69)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (70)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE
: LOCATION: (72)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala, Pro or Gly
: NAME/KEY: UNSURE
: LOCATION: (73)
: OTHER INFORMATION: Xaa is Pro or Asn
: NAME/KEY: UNSURE
: LOCATION: (74)
: OTHER INFORMATION: Xaa can be any amino acid or no amino acid, position 74-123
: NAME/KEY: UNSURE
: LOCATION: (124)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE
: LOCATION: (125)
: OTHER INFORMATION: Xaa is any amino acid
US-09-908-805B-59

Alignment Scores:
Pred. No.:      1.77e-06      Length:      128
Score:          123.00      Matches:      5
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    24.31%      Index:          0
DB:             10           Gaps:          0

US-09-880-711-328 (1-350) x US-09-908-805B-59 (1-128)

QY      185 NAGNAGTNTGCANN 171
      |||||:|||||:|||||
      DB      68 *****His*** 72

RESULT 15
US-09-796-448A-26
: Sequence 26, Application US/09796338A
: Patent No. US20020061522A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: TITLE OF INVENTION: 1983, 52481, 2398, 45449, 50289, AND
: FILE REFERENCE: 52872, NVEL G PROTEIN-COUPLED RECEPTORS AND USES THEREFOR
: CURRENT APPLICATION NUMBER: US/09/796.338A
: CURRENT FILING DATE: 2001-02-28
: PRIOR APPLICATION NUMBER: US 60/186,059
: PRIOR FILING DATE: 2000-02-29

```

```

: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 26
: LENGTH: 175
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: exemplary motif
: NAME/KEY: VARIANT
: LOCATION: 1-4, 6-53, 55-66, 68-137, 139-144, 146, 147, 150-170, 172,
: LOCATION: 173, 175
: OTHER INFORMATION: Xaa = any amino acid
: NAME/KEY: VARIANT
: LOCATION: 149
: OTHER INFORMATION: Xaa = any aromatic amino acid
: US-09-796-338A-26

```

```

Alignment Scores:
Pred. No.: 1.75e-06 Length: 175
Score: 123.00 Matches: 6
Percent Similarity: 85.71% Conservative: 0
Best Local Similarity: 85.71% Mismatches: 1
Query Match: 24.31% Indels: 0
DB: 10 Gaps: 0

```

US-09-880-711-328 (1-350) x US-09-796-338A-26 (1-175)

```

QY 190 TCCTGNAGNAGTTCANNNG 170
    |||||
DB 145 Cys*****Gly***** 151

```

Search completed: April 28, 2003, 15:39:57  
Job time: 22 secs